UCATIONAL and
YCHOLOGICAL MEASUREMENT

VOLUME FIVE, NUMBER TWO, SUMMER

*Philosophy and Practice of Personnel Selection.* HERBERT A. TOOPS ..... 95

*The Counseling Program of the Veterans Administration.* CARLOS E. WARD and GWENDOLEN SCHNEIDLER ..... 125

*Personality Traits Associated with Abilities. I. With Intelligence and Drawing Abilities.* RAYMOND B. CATTELL ... 131

*Factor Analysis of Occupational Aptitude Tests.* STAFF, Division of Occupational Analysis, War Manpower Commission ..... 147

*The Interests of Forest Service Men.* EDWARD K. STRONG, JR. 157

*A Group Testing Program for the Modern School.* WARREN G. FINDLEY ..... 173

*Replies of Psychologists to Several Questions on the Practical Value of Intelligence Tests.* ARTHUR KORNHAUSER ..... 181

*Measurement Abstracts* ..... 191

PRINTED IN THE UNITED STATES OF AMERICA
THE SCIENCE PRESS PRINTING COMPANY
LANCASTER, PENNSYLVANIA

# PHILOSOPHY AND PRACTICE OF PERSONNEL SELECTION

HERBERT A. TOOPS
Ohio State University[1]

By definition selection implies more candidates than jobs, a choosing of the most fit. In boom years and wars selection wanes; it waxes in depressions and peace. Following the war it will become important again.

It has become obvious that much of our material progress is due, on the one hand, to a very few expert people who are able to invent such things as B-29's, radar, dehydration, and penicillin; and equally, on the other hand, to a multitude of Joe's, Bill's and Sally's whose skill of hand, keenness of eye and sureness of touch, in small things, just as surely is an expertness of its own. Some kinds of people do each of these respective kinds of work better than others. Subdivide and specialize industry as much as you will and still there will be more work for each of these kinds of people to do besides all the more supplying work for a third class of experts, the managers, the Henry Ford's, the Henry J. Kaiser's, the J. F. Lincoln's, and others of lesser publicity and prominence. Expertness is important in all these realms.

Selection is both positive and negative. When looking for traits that are rare—in consequence of which we pay well for them—we wish to include as many as possible of the desired traits in one man; we look ideally for the one man of all men who most completely can fill the bill, the one man who includes in his make-up all the positive virtues. Of Tom, Dick, Joe, and Sally there are a myriad; hence they are paid chiefly for their time rather than for their pattern of abilities, and here we may seek only to *exclude* a certain few undesirable or nega-

[1] On leave with the National Roster of Scientific and Specialized Personnel.

tive traits such as dishonesty, unpunctuality, sickness, lethargy, and the like. We can afford—or so we often think—to hire them on an actuarial basis of "hire ten and fire two" since "not so much risk is involved in any individual mischoice" and "if they are not expert we can quickly and easily train them."

We also rely on the fact that what one workman fails to produce another may make up for by greater diligence; the shortcoming of one thus is offset by the superiority of another. In the creative, or inventive, and the managerial realms, on the other hand, the weakness of the one—whether superior or subordinate—is less compensable, and in fact is more likely to result in a weakness of both. Yet even here the idea has potential merit: Could one pick "superiors" to work in pairs—or in the general case in teams—so that the strengths of the one overcome the weaknesses of the other, and vice versa; the too conservative tendencies of the one curb the too radical tendencies of the other; the inventive tendencies of the one stimulate the productivity of the other, till in the end, over the years, both come, like long-wedded couples are supposed to do, to resemble each other highly in all the good and virtuous traits? In fine, what persons should work with what persons; what roommates room with what roommates, what persons be friends and pals with whom, and who should marry whom? We should consider such questions in terms not only of the traits they each possess but also of the traits they may wish to develop. This, of course, is a discipline mainly for the future. Its statistical difficulties are mainly a difficulty of notation. We have, so far, a dearth of studies along such lines. It involves questions not only of comparing profiles of persons in their present or cross-section aspects, but also of considering them in respect to their productivity and their probable future trend.

About the course of growth curves we are not so certain as we once were. Thanks partly to war, we are not such dyed-in-the-wool hereditarians as formerly. In war—which incidentally has come to America with distressing regularity over the decades of her existence, itself owed to war—the second and third generations which fight the succeeding wars have

in each succeeding crisis little precedent to guide them. Hence, in these latter days, now that scientists are recognized more for what they are capable of doing than merely as so many additional units in the supply of cannon-fodder, more and more innovations are being tried out. We discover that illiterates, not suffering from illiterate intellects, can be made "literate" in a matter of weeks; that most of the color-blind can be made color-seeing; that those deficient in the ability to see in the dark can be taught to develop "cat's eyes"; and that plodding farmers' sons may become great heroes of the air force. Do these not shake our faith in predestination pronouncements of a generation ago?

We sometimes lose sight of the fact that man, with his superior cerebrum, is the most adaptive mechanism there is, extremely sensitive to the world about him, and particularly to that portion of his environment marked off as "the people with whom he works and lives." So true is this that for every man who stumbles or falls, we, as practical psychologists, look for the woman in the case; for every divorce we suspect the spouse; for every turnover we suspect the foreman; and for every corporal's failure we suspect his sergeant. In a very real sense *the traits of one's fixed-relation associates thus are one's traits.* Enrolling such other-person's-data in parallel columns of the data book after the "personal" data, one has the makings of a two-curves profile involving this social relationship; amenable to all the techniques to which any data may be put, and with some inherent niceties, such as, for example, entering data by pairs (the paired associates' respective scores in a given trait) into a multiple-ratio regression equation. Compensation of traits would be revealed, perhaps, by opposite signs of the two associates' scores in a given trait.

We are becoming more concerned, too, about what are traits. If they are not inherited, or not so much inherited, what then? Do they boil down to some physiological tendency such as the fund of usable energy which the individual possesses, paralleled in turn by some simple index of amount of absorbability of vegetative tissues, or some set of more complex reasons geared up with the glands or the like?

In addition, we have come, on the one hand, through the work of the clinicians, who concern themselves with human motivation and more recently with human reformability, and, on the other hand, through the work of the statisticians who have concerned themselves with those individual cases which in correlation plots destroy the validity of the predictions, to question whether our erstwhile conception of personality as the added sum of its parts—or of its weighted parts—is the correct one. Everyone knows that some traits are more important than others—for particular purposes, or at particular times, especially at crises. This, then, is not to question the usefulness to society—which as a whole always works on the actuarial basis—of the concept of the ordinary multiple regression equation. The statistician will never be content, however, with validities in the fifties and sixties and an occasional higher one. He knows that the difference between an ideal and unattainable unity and his (always!) inferior index is in part due to, or is associated with, individuals who do not fit his simple hypothesis.

Add traits, either in numbers or of varied kinds, or both, and still the divergent individuals are almost as divergent as before; and the common observation is that despite all the work the multiple correlation coefficient is singularly unaffected. Can this mean that our form of regression equation is at fault implying in turn a wrong conception of the matter? As biometrists we are little concerned with the man who took to drink and went all to pieces when his wife died, the child who suddenly turned truant, or the genius who sometimes emerges from the two-room sod house on the prairie. [There are many more of such habitations (environments) than of mansions.] In general, we leave those "situations" to poets. Ought not our conception to be enlarged so that these cases, as well as those of normal school-children, of all ages, who hitherto customarily have been the subjects of our studies, are all subsumed under the same formula? There is little merit in explaining them away by naming them cases of shattered, peculiar, abnormal, or emerging personality.

Are there not indeed crucial traits which unless operative

at least to a minimum degree render all others fruitless? Is it conceivable, for example, that any American without a certain mathematical background (which few Americans possess but which many conceivably might acquire) could read Einstein no matter how intelligent, persevering, or able as a reader he is? This essential lacking, the product is zero, or virtually zero, no matter how favorable the scores in the remainder of the "causal variables."

In the statistical treatment both the units of measurement —in their mathematical aspects—and the form of the "regression equation" will surely be involved. Little is written on the matter now, but much more deserves to be and will be written about it in the years to come, particularly if a few mathematically capable or promising and ambitious people can be recruited to the psychological profession.

Finally, to round off our background of the matter, is it not clear that the cross-section aspect of a trait which we get as the result of a test or inventory (qualification form, questionnaire or interview, for example) is only *evidence,* rather than fact, only a straw in the wind of how the individual growth curve blows—or grows?

There are available but few growth studies of individuals in the several functions of physical growth, mental growth, and social growth. Those few suggest a functioning interdependence of such growths, predicated perhaps on a common fund of life-energy basic to the several growings resident within. They imply that the child is father to the man in more than years. It is useful then to look at the child if we would see the man, for what he is and yet may be. (That old dogs prefer to learn no new tricks is more of a truism than that old dogs can learn no new tricks.) And it is useful to keep track of him throughout life. Consequently a general growth study of the individual may be more revealing than any number and any quality of cross-section variables which ordinarily we may collect for selective purposes. Quasi-growth studies may be made from a consideration of the ages of the individual, computed for the dates of certain happenings, such as of his several job promotions, acquisitions of responsibility and title, and the like.

Present prognostic tests may be more so by name than by realization, particularly in the absence of long-range follow-up studies to convict us of our errors of "misappellation." The value of an "evidence" is measured not by its name but rather by its validity coefficient. Thus we see that the improvement of selection lies largely in the development of new techniques. Only a very few of its roots go back to the past. Our present practice accordingly may be found to be faulty, or be buoyed up unduly only by actuarial considerations which, partly or alone, save the day.

The last point is worth belaboring.

Two hundred men are let out of a modern plant, A, let us say, which is closing down on war production, on a Saturday. They have been trade-tested and it is known that 80% of them are "good workmen." The other 20% of "not so good" workmen are able to drive rivets, bolt-up nuts, and in general do any work for which they were hired, but with inferior speed and accuracy. In peacetime they would be "non-hirable" because too inefficient. Plant B is just starting up a new department. It advertises on Sunday for 100 men to show up at its employment office on Monday morning. At seven o'clock Monday morning let us assume that all of the recently discharged 200 are standing in line outside B's employment office door; also that in recent days the government has doubled its initial order with B, so that the word goes down to the employment office to "Hire them quickly so that we can get them to work. We can fire those who don't fit in. The government will pay the bill." So our obliging employment clerk counts the men in line, notes that there are 200, mentally calculates, since they want only 100, that "flipping a half dollar will do the job nicely and moreover will give every man an equal chance at a job." He makes a little speech to that effect and hires the men accordingly. And all are happy, even the men who get no job—since they had their chance—as though that were a good in itself!—save the father whose infant daughter badly needs at once an operation which only a job can pay for. (This is statistics and probability, not humanitarianism or social security.) So in fifteen minutes the impartial half-

dollar responds to "Heads we hire this man; tails we don't" in its expected manner and soon one-half of them, 100 in number, are at work. Silly? Perhaps! But what is the worst job that could possibly be done by our half-dollar under the circumstances? The answer evidently is: To hire all 40 of the "not so good" men. (Twenty per cent of 200 men equals forty men.) That would yield a selection efficiency of 60 per cent. The half-dollar would do the job as badly as that "only once in a million times." What is the best it could do? Obviously, to hire 100 of the 160 "good men," and reject all 40 of the "not so good" as well as 60 of the "good," with a resulting index of efficiency of 100 per cent. This too is a very improbable occurrence. What is the most probable result? Evidently to hire 80 good men and 20 not so good ones, resulting in 80 per cent of efficiency of placement. (The same figure which the trade-tests revealed.) If classification experts, placement clerks, manning tables, job families, morale officers, and all the rest, can beat this figure, the excess—above the 80 per cent—can be credited to their efforts; and if a lesser figure is obtained, their value then is all on the negative side of the ledger!

Let us carry this matter a little further back in time. We note, then, that much selection has occurred before the 200 men started for the doors of our employment clerk early this morning. This may conveniently be said to consist of two types: self-selection and pre-selection.

The self-selection consists in such facts as that only men able to walk showed up at the office; only men able to read or possessed of relatives and friends who could read; that no coffee-tasters, goldbeaters, astronomers, mind readers, college professors or tight-rope walkers applied—indeed only people who wanted work, *this* work which they judge, rightly or wrongly, is like what they have done before or which they hope they may be able to qualify for.

The pre-selection consists in the effect which the ad, mainly perhaps, had on the potential applicants. If the ad says "gentiles only," few Jews will have the temerity to apply. If it says "white or colored," some colored who otherwise would

be hesitant may decide to apply. If it says "men only," what woman will apply? If it says "$10,000 and expenses" what $200-a-month clerk will apply? Indeed selection is benignly affected, all unkownst, by a myriad of such considerations which cost little or nothing of effort or thought—and which, in passing, may have just as little to do with the efficiency on the job of those hired, the lucky ones we would call them a mere four years ago, and perhaps still might do so with almost as much justice.

That the patient gets well without doctor is counted a miracle and that he gets well in spite of the doctor is never counted a blessing! If the doctor is called, all the world including the patient is satisfied that a great good has been achieved! But we—those of us appointed to note and to better such things—cannot be satisfied merely because our clients are contented! We find it necessary, useful, and right to look into the selection process with a critical eye. One fruitful way of doing this is to look into the implied statistics behind each of the current or possible modes of selection. It might be well, however, to delay that consideration for a little in order to ponder another matter or two and particularly one strictly statistical principle which grows out of the successive hurdles method of selection.[2] This method implies that only the passers of a given examination are allowed to take the subsequent qualifying examination; and evidently this test may also be a statistical one, a sieving process applied to the data of the several individuals on a common trait-profile recorded in their record cards; only those who qualify on the first trait being allowed to be considered for passing the next test, the next trait, the passing point of which comprises the next hurdle. In high-jumping contests we do not necessarily rank everyone exactly in the order of their jumping ability by ruling out those who do not (we did not say "cannot") jump the bamboo at three feet, a *difficulty* measure we say, forgetful of other facts such as that the motivation of severe competition often determines whether a record is broken. In "tests" of selection, not the difficulty, but rather the validity, of the test is the impor-

[2] Toops, Herbert A. "The Successive Hurdles Method." *The Personnel Journal* 11 (1932), 216–218.

tant consideration. We get more or fewer jumpers merely by altering the height of the bar, by lowering or raising the passing point of the test. The all important consideration in selection is what kind of persons got over the bar—passed the test—and are allowed to take, or are subjected to, the next test. Accordingly if a trait of low validity is used for the first sieving any successive sieving of humans in selection, however fine, cannot undo the effect of having let most of the better individuals go into the discard. Still more concretely, the ends of all selection are at least two in number:

1. To assure a quality of those selected which shall be materially above that of those rejected.

2. To reduce a larger number of "applicants" to a smaller number of those recommended or hired.

The former objective, of course, is the important one. It follows then that in attaining the second end, that of reducing the number of candidates, it is important to keep the merit of the retainees as high as possible and to maintain it there as long as possible in the process of reducing the number of applicants. Briefly, it finally amounts to this: *that the ideal selection is achieved when the successive hurdles or tests are applied in descending order of their validity coefficients and that the application of additional tests is stopped arbitrarily whenever the number of applicants decreases beyond a given minimum.*[3] If the tests have no validity at all, as is true of our half-dollar above, then the order of application of the hurdles is immaterial. All that the tests accomplish in this latter case is:

1. A considerable reduction in numbers by reason of the application of the *n* hurdles.[4]

---

[3] If, regardless of the number of persons who pass all the hurdles, it is decided that all the hurdles (or tests each with their several passing points) shall be applied, willy nilly, the order of application of the tests makes no difference; the same individuals will respond to the total process in every permutation of order of application of the selective variables. We assume that the passers of two tests will be more able than the passers of one only; but this expectation follows, the first test having positive validity, only if the second test has a validity higher than zero; and it can actually destroy some of the validity obtained if the second is of negative validity. The validity of two can be less than the validity of the best one alone also when in weighting methods of selection we "overweight" the poorer test beyond its multiple regression β-weight. It is by no means axiomatic that "two tests are always better than one."

[4] Five such successive hurdles, each eliminating 50 per cent of the previous retainees will reduce a field of 3200 applicants to 100. Two more, or seven in all, will reduce it to 25.

2. A building up, in the minds of the applicants, of a belief in the fairness of the tests. [A feeling which unfortunately can be, and very probably will be, almost as great for low validity, or zero validity (chance) tests, as for highly valid ones.]

So long as they were believed in, for example, tests of memorization of Confucius were "good tests" in China, but any statistician could tell you that the character of the civil servants of China under such a system could be no better than that of the average person able to master this first hurdle to political preferment. A good rote memory obviously is a valuable asset of a statistical coder but surely is not all important for a high civil officer—or at least most civil officers—of the state. This settles at one stroke any contention that a uniform pattern of selective traits—at least of those presently available—could be equally valid for all purposes, for all occupations for instance. Traits are important—or so we now consider—in terms of their weights in a multiple regression equation. Alter our conception of the regression equation and some new measure of the relative importance of the traits results. At present the traits employed by sociologists, census-takers, and newspapers to typify the social-composition of the individuals of a social body may have zero or even negative validity—depending on the extent of their correlation with an independent criterion of ability in the occupation or undertaking in question—and yet may serve just as well as any more valid set of traits to quickly reduce the number of applicants to a number feasible for "more detailed examination (usually an interview) and consideration before appointment."

Where selection of equals is at stake and competition for jobs prevails, clearly justice requires that each person be given an equal chance of being selected. Of course this principle does not apply to unequals, for the better man, other things equal, should always be chosen. And if things are not equal but are almost equal, then there surely is some social gain if the renouncement can be voluntary. Not every teacher, for example—who incidentally happens to have a husband—can see any justice in giving up her job to an inferior, equally good, or even slightly better teacher who happens to be both husband-

less and jobless. No obvious improvement—and possibly some loss—in the condition of the taught is foreseen as the result of such replacements. We are only in process of establishing the morality that one's right to a job and to advancement is independent of his social and economic status.[5]

The severity of the critical scores employed in the successive tests has a bearing not only on the amount and the quickness of elimination but also upon the validity of the sortings. To eliminate no one by a highly valid test is to eliminate entirely its validity as a selective agent, to cancel it out as if it had never been given. We are unable to state this effect in other than the very general statement[6] that the more valid tests should have the more rigorous critical scores if there is any variation of standard in the several hurdles.

There was an intimation above that under some circumstances there was merit in an actuarial hiring of all who apply and in letting the test of the job decide who should be retained. On a cost-plus economy this is always feasible particularly if:

1. The percentage of applicants who are above the minimal "acceptable point" of competence is high. (In our example above, if only 50 per cent of our applicants instead of 80 per cent were "competent" the impartial half-dollar *could* give us no satisfactory selectees at all. With the number of competents less than 50 per cent, the probability of such an occurrence is higher and higher as the percentage index descends.)

2. There is some highly efficient method of quickly weeding out the incompetents on the basis of their performance subsequent to selection.

These considerations yield us our first potential or actual selection method.

### *The Test of The Job*

If we measure day by day the rate of learning (or "progress") of beginners, say, by means of adequate work records,

[5] The war has given us a new morality about the matter of one firm or one service gutting the market for good men, merely because it happens to be first, or have the most money, and the like.

[6] An important initial attack on this problem has recently been made by Richardson, M. W. "The Interpretation of a Test Validity Coefficient in Terms of Increased Efficiency of a Selected Group of Personnel." *Psychometrika*, IX (1944), 245–248.

both a prognostic test and a trade test may be unnecessary. Relative indices of progress on the job may then be all the test we need. This is true because there is some warrant for believing that differential capacities yield individual progress curves, with distinctly different rates of growth, which differentiate at an early date. The rank orders of the growth curves of the apprentices at an early date thus approximate the final orders of competence. Our technical problem then is to shorten the tryout period as much as possible.

Let us assume that all of the new workers in a certain factory are put through a sequence of jobs, first on the drill press, then on the shaper, then on the planer, etc. It is clear, then, that the number of hours that it takes candidate A to complete a standard drill press assignment as compared with the hours required by candidate B is some indication of which of the two is presently more competent thereon; perhaps some slight indication, also, of which of the two gives more promise—if both were equally "untaught" at the beginning—of future usefulness in that particular drill press department; that is to say, possesses more "drill press aptitude." At the end of the ensuing shaper operations, which it is assumed follow in all cases upon the drill press operations, the *cumulative* number of hours required for completing both assignments gives a better measure of the "all-around" mechanical capability of the candidates than the previously mentioned shorter "test" dealing with only a single (somewhat more specialized) ability. Accordingly, the greater the number of pertinent experiences included in the testing the more is revealed the "all-around" mechanical ingenuity, or general mechanical ability or adaptability or learning-power, of the (apprentice) learners. The question now becomes a statistical one: "What is the earliest period at which the cumulated competence scores can be made to correlate to at least a minimum limit, say .90, with ultimate competence?" The end sought is to minimize the time, work and money necessary to make a valid decision of whom to keep (because worthy, probably, of promotion) and whom to discharge or transfer (because of inadequate aptitude for the work). The order of presentation of the work experiences, drill press-shaper-

planer must have been fixed upon, of course, and be uniform for a particular entering group of apprentices, but may be altered on subsequent groups to place the most valid shops (those correlating highest with the sum of them all) into the earliest positions in the try-out experiences, as is necessitated by the successive hurdles method. On our initial experiment all must complete all assignments to arrive at an over-all criterion score for each. Employing the L-technique one now may easily ascertain how highly the first shop correlates with the sum of the success scores involving them all; how high the score of success in the first two shops combined correlates with the total success; how high the combination of the first three, and so on. The correlation will become high at an early stage, particularly after the shops once have been arranged in an order of decreasing correlation of the several shops with the total success on all the shops.

An alternative, and preferable, method is to place in first or accepted position in our upbuilding composite job that task which correlates best with the sum of them all; then, in second place, that one of the remaining which raises the correlation most; then that of the remaining tasks which raises the correlation next most and so on. The eventual order is "that order which maximizes the prediction of the criterion with a minimum of tests." We have here a choice of utilizing either the multiple-ratio method[7] or the L-technique, the latter being preferable wherever simplicity is important. The former does not require the computation of all the inter-correlation coefficients, which is an advantage if the number of elements to be combined is large. If all the intercorrelations are available then the Wherry-Doolittle technique is appropriate and has merit for the purpose of selecting a minimum of "shops" to correlate maximally with their total.[8] This simultaneously minimizes the time, and optimalizes the order of presentation, of the accepted shop experiences as a test of machine-shop aptitude.

[7] Toops, Herbert A. "The L-Technique." *Psychometrika,* VI (1941), 249–266. See also, Toops, Herbert A. "A Self-checking Technique for Shortening a Test." *O. C. A. Bull., No. 92,* Jan. 30, 1934, 2027–2040.

[8] Stead, Shartle, *et al. Occupational Counseling Techniques.* New York: American Book Co., 1940, pp. 245–250.

In all such methods some regression of the validity coefficients of the selected scale, on subsequent tryouts, is to be expected. The point at which the thus obtained correlation, by whatever method, first reaches .90 or .95 say, will mark off both what shops it may be necessary to give to all candidates and also the desirable order of their presentation. The individual's score on the accepted composite will reveal substantially whether a given individual does or does not possess sufficient "aptitude" to be allowed to continue. Where there are necessary sequences in learning skills—of which there are few in either school or industry—the statistically dictated optimal order of presentation cannot of course be employed.

Let us suppose that as a result of an adequate experiment in which a number of beginners have carefully been watched through to ultimate competence, it has been decided that the number of hours needed to complete the first five (thus reordered) operations is a highly valid measure for prognosticating ultimate and "all-around" success. It is clear, then, that if we build five successive sets of norms of cumulative progress as of the ends of these five several operations the "normated" individual performances obtained from the job itself on the five successive occasions of completion of an additional project are progressively more and more indicative of the ultimate worth or lack of merit of the individual apprentice of concern. For solution of our statistical problem no test, then, other than the test of the job itself, is necessary. Then if industrial agencies (as of wartime) demand a decision which may be made quickly with only fair accuracy it may be made after the first operation, with somewhat better validity after the second and so on, the validity improving with the length of the tryout.[9] By the steps outlined we have discovered the minimal number of operations (involving their identity and sequence) which, with the least possible wastage of supervisory time and overhead cost, enables us to arrange the candidates

[9] As in sermons, where it is said that most souls are saved in the first fifteen minutes, so with tests. Tests or hurdles beyond the first highly valid one add mostly to the reliability of the scale—and so to the "justness" of the exclusions, when considered from the view of the excludee or potential excludee—but add little to the validity of the scale.

substantially in the order of their relative excellence—centile ranks, perhaps—for retention (and promotion). The above applies when even the poorest man turns out a salable product.

But the problem is not quite so simple as the above statements would imply. "A man is worth the selling price of what he adds to the product less a reasonable profit" is an axiom in economics perhaps but vocational psychology is not quite so naive as that. Observation shows that the worth of a man is measured adequately neither by the wage paid him nor by the quantity of work he turns out. Both are only *symptoms* of his worth or value. The slow-learning but ultimately fairly capable employee may be more of a "success," from the employer's viewpoint, than the quick-learning person who is the delight of the teacher or foreman (the typical factory teacher). This statement follows if the employer values highly traits other than productivity, matters such as punctuality, dependability, or gentility. No series of measures can ever completely describe the man. The individual is what he is in the measures employed; he is something else when more or other kinds of measures or tests or aspects of worth are applied. The employer then may indulge his *preferences* (or prejudices) between such "types" as a "slow but accurate" individual and a "rapid but somewhat inaccurate individual." Concretely, let us assume that the first of two persons rates 40 for speed and 60 for accuracy. If the arithmetical units are comparable, so that they may legitimately be averaged, the first individual then averages out to 50. Let us suppose that a second individual rates 60 in speed and 40 in accuracy. His average also is 50. But the two individuals, thus seeming alike, each having an average of 50, are quite different in the employer's eyes. The first will not get so much work done, but he has the merit that he will not destroy so much raw material; the quality of work produced will be better; and he probably will not require so much supervision and correction of his work. Such considerations are "values" to most employers. Inquiry on the part of the U. S. Civil Service Commission, for instance, has revealed that business men are quite content with a lower typing speed in a secretary than that set as the required gradua-

tion speed by a majority of stenographic schools provided only that they can acquire people who do not make so many mistakes (even if they don't get so much done) which need to be corrected. In line with this, employers rate highly the ability to spell (to correct their boss' mistakes).

The above homely example indicates that success is not a "unitary" concept; it cannot be measured in one dimension but can be only more or less adequately portrayed by a profile, a series of scores in a number of variables, measuring ideally different "dimensions" of the applicant. Without tests, can we still measure the individual for selection in terms of a profile, to the end that the profiles of different candidates may be compared and a choice be made? By making the permissible assumption that in some occupations at least a man's past portends his future—which is a safe assumption in fields such as teaching, management, and the professions generally, wherever occupational growth requires considerable time and is continuous—this is readily done by a technique previously published by the writer,[10] and may be briefly summarized as to its principles or elements:

*The Test of Accumulated Evidences*

1. An extensive qualifications blank and specific recommendations blanks received from former employers or teachers provide a great deal of concrete evidence, from various angles, as to an individual's success to date in his occupation and his adjudged promise therein.

2. By the aid of official evaluators of such evidence, who function as a standing jury in that capacity year after year, the lines of evidence are telescoped into quantitative scores, of some objectivity and validity, recorded as some fourteen (present practice) aspects of ability deemed important for the job in question: intelligence, research ability, scholarship in the specialty, general scholarship, social intelligence, managerial and executive ability, special skills useful in the occupation and the like.

[10] Toops, Herbert A. "The Selection of Graduate Assistants." *The Personnel Journal*, VI (1928), 457–472.

3. The resulting scores, when plotted, become the candidate's profile just as truly as if one had test scores, social composition items, standard scores on a standardized progress schedule, and the like.[11] The subsequent evaluation may now be by any of the general methods given below. In the article referred to, the subsequent evaluation was by the summation-of-traits-scores method, which may or may not be best for one's purpose, as noted hereinafter.

Given, then, a profile of the individual, assuming all necessary corresponding statistical implications, there are a good baker's half-dozen different ways of selecting talent, that is, of reducing the number, while increasing (as an expectation) the competence, of the retainees. All of these in common assume that the pertinent traits (or at least the traits on which he selection is to be made) have been observed, preferably objectively measured, on each of the candidates and that the scores necessitated correspondingly are available.

1. *The summation-of-traits-scores method*

1.1 *Where the scores are added at gross-score weights of 1.* This is the method customarily used in most school examinations where the sum total of the person's merit is simply the arithmetical sum of all of the scores on the several sub-parts (sub-tests and items). This implies an addability of scores which seldom obtains. *This method,* contrary to popular belief, *does not weight the sub-parts equally, but instead weights each item, and each test, proportional to its standard deviation.* In many cases this may not be such a bad assumption, however, since the test which can produce the greatest spread of scores in general is the most valid.

1.2 *When the several scores of a given person on several traits or sub-traits of the total examination in succession are multiplied by a fixed series of arbitrary individual test gross-score weights.* This is the "weighting" method which is popular with civil service gen-

[11] Toops, Herbert A. "The Criterion." *Educational and Psychological Measurement,* IV (1944), 271–297.

erally. It is likely that civil service frequently has greatly erred in the weight which it felt it was attaching to the scores on the several sub-parts in question.[12] It is a safe guess that, the standard deviations abetting, many a man has been selected by civil service mainly on account of his handwriting or some other comparatively unimportant trait. Such "errors" can be avoided by changing the original scores to ranks, standard scores, *T*-scores, or other relative measures in which the standard deviation of the different traits becomes a constant. Standard scores have a standard deviation of 1 while rank scores have a standard deviation of $\sqrt{\frac{N-1}{12}}$; *T*-scores a $\sigma$ of 10, and so on, irrespective of the variable measured.

The weights may be ascertained by the bids procedure of the ensuing footnote[13] which standardizes the process of ascribing the weights and replaces the presumably inferior judgment of a single person by the possibly superior judgment of a group of "experts."

The weights may be obtained mathematically, a criterion being available, and thus have the superiority ascribable to least squares technique. It will be recalled, however, that a unitary criterion score may come from vastly different profiles of "success variables."

The most popular of selective methods fails then to predict well even criteria whose similar mode of compiling undoubtedly weights the scales at least slightly in the direction of high, rather than low, validity. It is perhaps the simplest concep-

---

[12] If $\beta_1$ be the importance actually assigned to standard scores, then $W_1 = \beta_1/\sigma_1$, where $W_1$ is the proper gross score weight to be employed in order that the trait shall be weighted with a true relative weight or importance of $\beta_1$. Conversely, if a gross score weight, $W_1$, be arbitrarily assigned to Test 1, the true relative weight or importance thus assigned is *not* $W_1$, but $\beta_1 = W_1\sigma_1$. Clearly, $\beta_1$ and $\beta_2$ are *proportional* to $W_1$ and $W_2$, or conversely, only when $\sigma_1 = \sigma_2$; or in general, only when the $\sigma$'s of the tests (of sub-parts) are equal. Standard scores, ranks, and *T*-scores among others, have this property.

Clearly also one may vary the length or difficulty of the examination in order to achieve approximately this end, a point of finesse worth further exploitation. If a test is lengthened to $n$ times its present length its new and enlarged standard deviation is $\sigma_n = \sigma\sqrt{n + n\ (n-1)\bar{r}_{11}}$ where $\bar{r}_{11}$ is its reliability coefficient.

[13] Toops, Herbert A. "The Selection of Graduate Assistants." *op. cit.*

tion of the organization of traits, merely the sum of the weighted scores, and is suspect for that reason alone—as is the I.Q.; for what else, pray, could one do in the latter situation but take a ratio! Man prefers simple laws; Nature, perhaps!

2. *The successive hurdles method*

This, referred to above as a principle, is the method which is popular with civil service in times when the supply of candidates greatly outruns the possible demand for placements. An examination is given in common to the entire list of candidates who have qualified on certain preliminary *minimum specifications* (health, freedom from arrest, etc.). A critical score is established above which all of those who "pass" the first examination are certified as entitled to take a second examination, while those below are denied further opportunity to qualify. Successive examinations eventually whittle down a large list of candidates to a few who are certified for appointment. The author[14] has shown that, when using this method, it is highly important that the tests be administered in the order of the most valid test first, the next most valid second, and so on down to the least valid last. The least valid and last test may then be practically a chance test. (It should not have a negative validity coefficient, of course, for that would reduce, instead of improve, the mean talent of the retainees.) Accordingly, if chance is to have a hand in the selecting of candidates who eventually are offered positions, clearly chance should operate only on a group of people who as a result of preceding screenings have a very high degree of capability. This condition easily is procurable by the sequence recommended. Clearly it were better that all the tests should be equally valid (and ideally intercorrelate zero), but since this is impossible the principle stated is obviously the correct one. From the viewpoint of the function performed it is clear that an individual always is thrown out by a single test, which in the nature of all tests has at least a degree of unreliability and a validity which may or may not be very pertinent, i.e., valid, to the particular job

[14] Toops, Herbert A. "Sifting Civil Service Applicants by the Successive Hurdles Method." *The Personnel Journal,* 11 (1932), 216–219.

which a given candidate is to fill.[15] This method had its origin during the depression when thousands applied for positions for which scores or at least only hundreds had applied before. It is the product of desperation! It is costly; and at best is not the method which one of free choice, cost being no consideration, would choose.

For many a candidate his destiny is decided by only one or two traits (the earliest tested) of his entire profile. The average competence of 100 candidates chosen by such methods —if the selective traits are valid—of course is high[16] when the tests are administered as above indicated, namely in a decreasing order of validity, and with the more valid tests producing the greater (and earlier) elimination of the applicants. This is the method employed by the Westinghouse Science Talent Search. Functionally, this method has much in common with the precise profile method below (q.v.). It has the merit that by setting a more severe standard the testing can be terminated briefly, thus minimizing its cost.

3. *The precise profile method*

In this method a certain number of traits are assumed to be highly (and equally) important to the extent that if an individual does not have precisely the skill-pattern, or profile, established as important (possibly by preferences of the prospective employer) he shall not be considered. The traits of a large number of individuals having been punched into Hollerith cards or Findex Cards, say, the subsequent sorting process yields all those persons who exactly fit the "selection profile." If too many candidates result, another trait, of little validity perhaps, may be added to the pattern to reduce still further the number of cases; or the subsequent selection may be subjective, based on "a detailed examination of the applicant's entire dossier." (This implies a positive validity co-

[15] Obviously the larger the roster and the fewer the referrals the greater the competence of the actual referrals, if the given selective method has positive validity.

[16] The normal expectation in lengthening a test is that this will increase greatly its reliability and increase slightly its validity. One of the "easy" ways to improve the validity of a test is to *shorten* it by eliminating the least valuable items. Tests of 100 or more items often will be found to have items of negative validity in them.

efficient of the examiners' judgments.) Only those who qualify on *all* the traits are acceptable. The machines are blind; they know no "humanitarian considerations." The method, as outlined, presupposes that all the traits are applied without fail to the candidates; hence the order of sorting of the traits (if one was employed) can have no influence on the competence of the selectees. The method presupposes that the needed numbers of candidates have been measured on all traits, and the task is that of picking the "most likely" candidate for placement.

If Findex cards, and quantitative variables, or qualitative ones possessing intrinsic quantitative characteristics are employed the chief fault of the method may be somewhat mitigated by slotting[17] on the "or more" basis in order to pick out, automatically, candidates who have *at least* the minimum standard, *or more,* on each trait. Thus, any person who is 9 in a trait is also slotted 8, 7, 6, 5, 4, 3, 2, 1, and 0; while any person who is 10 is also slotted 9, 8, 7, 6, 5, 4, 3, 2, 1, and 0. If, then, one sets, say, 7 as a minimum in a selective trait, all persons who are 8 or above and 7 or above also will respond to the selector rods of the mechanism, as well as those persons whose score is precisely 7. In the above case, if one sets 10 as his minimally acceptable score, only the second candidate will respond to the rodding; but if he sets 9, or 8, or 7, say, both candidates' cards will respond because both are equal to or are greater than these several limits, respectively. If not enough candidates show up, one may then reduce his minimum to 6, say, and find additional cases who are 6 on the test in question to "fill one's quota," on the assumption that one wants to pass some definite per cent of all the applicants on this one trait.

If the entire pattern of traits is specified at the outset and is applied to all the cases and this results in turning up several candidates for each one to be hired the final selection may then be made on the basis of "personality" traits or even on other considerations having zero (but not negative) validity. If too

[17] If Hollerith cards are employed, the same end may be achieved by sorting the several traits *in succession,* picking up for subsequent sorting not only all those "passing" but also those of higher standing in the trait in question.

many candidates show up, one may raise his minimum requirement in the trait or traits in which it is known or is judged that such raising is likely to be most profitable (namely the trait or traits of greatest validity.)

Some of the traits may be expressed as *unvarying standards;* that is to say the particular category in the trait in question is the only one acceptable. Thus, one may insist on having a male teacher, females being totally unacceptable; or an unmarried teacher; or a Protestant teacher; or any compound characterization of these,[18] e.g., a married-man, or a married-man-who-is-also-a-Protestant. With respect to traits which are strictly quantitative, such as intelligence, scholastic average, height, weight, etc., usually anyone *above* a certain minimum is deemed acceptable. Sometimes, however, an upper limit is established as well. That is to say, policemen must be at least 6 feet, 0 inches tall and not more than 6 feet, 4 inches. (This would imply the existence of a curvilinear relationship of the traits upon the independent measure (criterion measure) of job success, for which selection is being carried on.) If Hollerith equipment is employed similar conditions may be made to prevail either by multiple-punching one-column traits in similar fashion (this precluding tabulation) or, better, by the selection of particular packs of cards for *successive* sortings. The multiple-sorting head device is essentially the exact profile method.

This "or more"-variation of the method resembles the popular method used by housewives in picking maids, by police captains and army recruiting sergeants in picking recruits, by employers in hiring workmen, and the like. It is frequently used by placement bureaus to place individuals, the card being removed from the "live file" just as soon as a given person is placed. It is the method employed by the National Roster of Scientific and Specialized Personnel, with the variation that an already employed person, if an "essential specialist," may be referred to a prospective employer whose priorities warrant.

[18] Only by an adequate follow-up, *when all are* hired (as in wartime) of each such compound type can we ascertain the relative merit of such qualitative qualifications. It is not impossible that some of these are on the negative side of the ledger. Humanity undoubtedly pays through the nose for its hiring prejudices.

Clearly the method presupposes a great supply of persons for a limited number of placements. Its essential difference from the successive hurdles method if the or-more punching or selection is employed, is a distinction without a difference. The fact that all the possible candidates have had all the "tests" (e.g., have answered all the questions of all application blanks and supplied all required credentials) and that the entire pattern is employed makes not a whit of difference. A too-low score on only *one* trait eliminates him sooner or later, but just as surely. If the result of the compound dragnet (applying all the hurdles at once instead of in succession) is unsatisfactory, the test profile may be changed until a "likely" candidate comes to hand.

Its chief merit is its potential flexibility. The housewife, so inclined, for example, may interview all who answer her ad and if many respond raise her pattern of requirements to "get more for her money."

When the traits are set by employers ignorant of trait implications the result obviously can become little more than a means of securing a random selection among the available candidates and of indulging the whims and prejudices of the requisitioner. The best to be hoped for is that any such pattern in general has "some pertinence" to the jobs in question.

Its great fault is that the candidate who is generally very superior but lower than the minimum in only one trait—who is 100 centile, say, in all traits but one in which he is, say only one centile below the established minimum—will fail to be selected. And since the selection is blind—i.e., the operator sees only the cards of those who respond to the selection, and is blissfully unaware of the "all but" cases—the conspicuous failures of the method—its shortcomings are never noted.[19] The Findex is limited to selection where only such numbers of cards (individuals) are involved as can be manipulated for hand sorting. Larger numbers could be handled by having

[19] This presumably might not be the case with Keysort or Speedsort cards since the "almost" cards would have a single solid card across a continuous channel of holes if the card in question chances to be in the midst of the "all but one operation to be performed" cards.

the cards filed in pre-sorted compound-breakdowns, which to that extent may defeat its own end.

This might be called the fastidious method of selection. It reminds us of the diner who to quench his thirst must have wine; and not only wine but Chablis; and not only Chablis but Chablis of a particular vintage, 1927, say; and who moreover is willing to pay any price, including that of not drinking, in order to appease his whim! Unless the available supply of applicants is large the precise pattern may locate no one. This, however, oftentimes is the consideration which makes the method appeal to the politician who would nullify civil service; since with a not too large roster an employer given that option very easily may set up a pattern of traits that none but the intended recipient of the proffered position will be able to meet.

In practical application the practice is to establish *by judgment* an ideal pattern; and then to investigate the available supply of applicants by the Findex or by the more recently perfected Multiple Pattern Sorting device of the Hollerith machine, an attachment for the sorter which will sort out all cards of any given pattern not exceeding ten contiguous columns of the eighty-column Hollerith card. Then if the initial sort yields no applicant, or not enough applicants, the pattern may be altered to a less ambitious one, whereupon more applicants will come to hand. In theory the "less important" qualifications are dropped, but who, in the absence of appropriate regression equations can say which are the "less important" qualifications!

Where prospective employers set the patterns, if they could be induced to establish *an order of decreasing desirability* for sorting the cards, and if the number of desired referrals could be always specified also, or else be predeterminable by formula *from the number* of workers requisitioned, then the sorter would have a freedom which permits of lowering or raising the standard of competence in order best to fill the requisition. Even here "stereotyping" of orders may reduce greatly the chances of referral of a man who is substantially good but not in exact agreement with the stereotype. The short man who on other scores would be a superlative policeman is out of luck

so long as more than enough 6-footers respond to that universal requirement! And this despite the fact that modern automatics are supposed, for some purposes, to make all statures even!

4. *The minimum divergence from the desired profile method*

Having an "ideal profile" in mind the candidate's profile of test scores can be matched against this profile and the candidate then can be hired by reason of an index of his agreement or lack of agreement with the ideal profile in question, say "that of the job." Probably one of the very best measures of the extent of such agreement is Sagebeer's index.[20] This index has a value 0 when the profile of the individual exactly matches the profile of the individual for which the selector is looking, while larger values of the index indicate greater discrepancies between the candidate's profile and the ideal. The method obviously assumes that the traits are *measured* scores and are *comparable*. Possibly the scores should be standard scores, thus making the assumption that equal standard scores in two or more traits are equal. The method will be most meaningful, perhaps, in large groups of applicants, all tested on quantitatively scored tests which are highly reliable and as nearly unique as possible. Possibly it applies better where aptitude rather than achievement is the basis of the selection. *The ideal of occupational selection, as of any other type for that matter, is that all applicants, not merely those who happen to be labeled by the particular magic name of an occupation, for example, should be scanned for selection.* The above index undoubtedly would be too laborious for this purpose without some machine method of solving the formula. This method, employing the index, has possibly never been formally used or used very much because of its newness. The method presup-

[20] Sagebeer's index is:

$$I = \sqrt{(X_{1_I} - X_1)^2 + (X_{2_I} - X_2)^2 + \ldots},$$

where

$X_{1_I}$ is the *ideal* profile score in trait 1, that established authoritatively by job analysis.

$X_1$ is the candidate's score in trait 1.

The correlation coefficient between the scores of the candidate's profile and the ideal profile would be an alternative index for the purpose at hand.

poses that every person has been compared with the ideal profile and a measure of his disparity therefrom has been obtained. Assuming acceptable *comparable units* for the several tests, this is possibly the most ideal of all methods, for general selective purposes. It tends to prevent an individual who is very lopsided from being chosen and yet allows deviations from the range of ideal scores to be considered—although at a (slight but appropriate) disadvantage.

5. *The predominant or outstanding merit method*

The essence of this method consists of the following operations:

5.1 On a sheet of cross-section paper, in Column 1 are entered the names of the several candidates.

5.2 The headings of the several columns are labeled with the various traits on which the candidates are measured.

5.3 Into the compartment is written the most significant statement (preferably quantitative) of the candidate's standing in the trait in question. This is done for all traits and all persons in turn. This then is the primary data table, or the Hollerith tabulator reproduction thereof.

5.4 Considering now trait 1 only, one may encircle in column 2 those statements which represent the *most meritorious* degrees of the trait desired in the candidate. [In general these will be the largest scores; in some cases, such as errors or time, they will be the smallest scores; and in case of curvilinear relationship they will be the scores that indicate greatest fitness for the job, that is to say, the scores closest to the apex of the parabola or curvilinear relationship line between the trait in question ($X$) and probable job success ($Y$).] Possibly the ten per cent or so of highest scores are so encircled. The number so to be encircled will depend both on the number of candidates and on the number of traits.

5.5 The same now is done for the remainder of the traits.

5.6 One may now either count and record the number of circles for each person (that is, the number of circles *per row*) or, as an alternative, may count up, for each person, the sum of the *weights* of the traits appropriate to the columns in which a given person possesses circles.[21]

The end achieved by this method is to give unusual weight in selection to those persons who possess *unusual excellence in more than one trait*. In other words, it is rather the antithesis of the method in which we set up minimal test scores and are satisfied with any one who, so far as the single trait is concerned, is at least above that point. By this method we assure ourselves that the appointee surely will have several strengths—if any candidates possessing such are available—although he also *may* have some very fundamental weaknesses; that is to say, he may be lopsided, but nevertheless he will be at least a "near-genius" in some respects. The method presupposes a great scarcity of the kind of talent (genius) desired and implies a desire to locate the "best" there is available.

This method implicitly, rather than formally, is assumed in case one goes out "combing the world" for unusual talent of any kind. Unusual talent is so scarce and withal comes so dearly that if one can get a near-genius in several traits, one generally is quite willing to overlook a number of minor or even a few pretty serious faults or defects. The adaptability possible in a *specialized industry* often makes it particularly feasible to place an individual where his defects, from the viewpoint of the job at which he has to work, will be practically no handicap at all.

One may vary this technique by encircling instead all scores of defects which one *wants on all accounts to avoid* and then consider for appointment him who has the fewest defects. This, presumably, formally approximates the procedure ordinarily used in choosing diplomats and other public relations

[21] The weights would be determined by the bids-procedure, presumably. If the β's appear on a T-square blade, with the data table tacked in alignment to a drawing board, the β's of the blade opposite the encirclements only are added. The T-square is a convenient device for bearing the marginal β's into the body of the table.

appointees where freedom from public offense is or may be a greater virtue than positive capability.

Still another variation would be to encircle the strengths and then to box-in all weaknesses, the score for a person being the number of his circles minus the number of his rectangles.[22] Such a method would require a larger number of registrants than either of the last two methods to produce the same number of selectees. It has the advantage that those selected have high basic strengths and at most only a very few fundamental weaknesses.

Other methods are no doubt possible. The above, however, comprise those which readily came to mind.

### *Comparison of Methods*

A careful inspection of the above systems will reveal that there is no one method which at all times and in all places is superior. Each performs somewhat different functions and has somewhat different emphases. To ask which of the methods is better, accordingly, is analogous to asking the question, "Which is the better container for apples: boxes, bags, or baskets?" The only reasonable answer is "It depends on times, places, and circumstances."

In the practical situation the matter of motivation of the potential employee is a portion of the total value to be placed upon a method, as well as the excellence of the sieving which it produces. It may or may not be wise, for example, for a candidate to get the notion that he is "the one person in all the world able to fill this job," even though, in a certain very real sense, this may be true. Human experience is more in accord with the sentiment that there are many people who could do a given job; and conversely, that for every person there are many jobs in which he might do at least minimally well. In view of the fact that we know so little about the functioning of traits, let alone their organization, causation, and growth, it seems safe to say that the summation of the weighted scores method, perhaps in the general case, is closer to the realities of "human experience" than most of the

[22] Or a similar formula in terms of β's.

other methods cited, with the exception that when one is looking for unusual talent (with the expectation of hiring only one or two individuals at most from "among all the world") then the predominant or outstanding merit method is obviously superior. When employing the latter method if one is going to hire a very large proportion of all of the available candidates, then clearly one wants to be sure to refrain from getting any large aggregate of "weaknesses" and the minimum trait method is good. In this case it is understood that it is the production of the group rather than of a particular employee which determines the personnel bank-balance of the selector. If one is going to hire quite a few candidates but still a very limited number compared with the total number of applicants, the successive hurdles method will be economical and withal effective, providing the method is carried out in the approved fashion outlined above; namely, the administration of the most valid trait first, then of the next most valid, and so on down to the least valid last. When using this method Ruml's rank tangential coefficient[23] may be of aid in establishing the "critical scores." (Critical scores are in none too good repute among psychologists today. The concept grew up in a day when it was believed that failures were due to one "type" of humanity and successes to another, and that accordingly, if only one built the right kind of a test, he would secure a bimodal distribution. If only this were not untrue, it would be a very definite concept in view of the fact that it would be easy, at least comparatively, to determine what point should be set in order that the overlapping of distribution *A* on *B*, should be a minimum.) Where the numbers of cards to be sorted is "small" and the categories of traits are few in number, Keysort or Speedsort has several advantages, namely, that (1) a sorting needle is the only paraphernalia required and (2) some capital presumably may be made of the visual channels on the edges of the cards as successive sorts—preferably in a descending order of the weights or importances of the traits—are made. In this latter case an

[23] Ruml, Beardsley. "The Reliability of Mental Tests in the Division of an Academic Group." *Psychological Monographs,* Vol. 24, No. 4, Whole No. 105, 1917, p. 59 ff.

unbroken card across channels, the sorts being nearly completed, would indicate the case of a person below the minimum (how much can quickly be inspected on the card in question) on the category in question under observation.

There probably are other methods of selection. If so, it is hoped that this paper may encourage others to discuss the remaining methods and to compare their advantages and shortcomings with the methods herein outlined. Selection throughout the world—particularly in areas where so much leadership personnel has been destroyed by war—will be more important in the days to come than in any recent historical period.

# THE COUNSELING PROGRAM OF THE VETERANS ADMINISTRATION

CARLOS E. WARD AND GWENDOLEN SCHNEIDLER
Vocational Rehabilitation and Education Service,
Veterans Administration

THE Veterans Administration is preparing to make the services of qualified counselors available to veterans throughout the nation. Counseling is being offered to all veterans who are eligible for vocational rehabilitation under Public Law 16, 78th Congress, or for education or training under Title II of Public Law 346, 78th Congress, and it is evident that a very large percentage of the veterans of World War II will be eligible for the benefits of these two laws which are being administered by the Veterans Administration.

Public Law 16, 78th Congress, approved March 28, 1943, may be regarded as a Disabled Veterans Vocational Rehabilitation Act, since its principal purpose is to provide vocational rehabilitation to overcome the handicap of disabilities which were incurred as a result of service in the armed forces during the period from September 16, 1940, to the end of the present war. Vocational rehabilitation will usually be attained through training provided for each veteran to fit him for employment consistent with the degree of disablement and suitable to restore his employability.

Public Law 346, 78th Congress, approved June 22, 1944, the correct title of which is "The Servicemen's Readjustment Act of 1944," became so widely known as the "G. I. Bill of Rights" prior to its enactment by Congress that many people still speak of it as the "G. I. Bill." Title II of this Act provides that veterans who meet certain eligibility requirements and whose education was impeded, delayed, interrupted, or interfered with by reason of entrance into service, or who desire a

refresher or retraining course, may be given education or training courses at approved institutions of their choice within certain time limitations prescribed by the Act. The veteran's length of service is an important factor in determining the duration of the education or training to which he may be entitled. Comparatively few veterans of World War II would fail to qualify for at least one year's education or training under Title II, if they should choose to apply.

When the fact is considered that complete counseling service, including modern techniques of testing and the use of compendia of systematized occupational information, is absolutely essential to the selection of an employment objective and of a training course suitable to the disabled veteran in need of vocational rehabilitation, it is evident that this alone would require an extensive Veterans Administration counseling program. But when to this is added the Veterans Administration's responsibility for counseling all eligible veterans who request educational and vocational guidance in connection with their applications for education or training under Title II of The Servicemen's Readjustment Act, then the magnitude of the task which this agency faces takes on new proportions.

While "educational and vocational guidance" is the term used in Public Law 346 to designate this service, the Veterans Administration has developed its counseling program with most careful regard to the fact that veterans will have greater assurance of achieving their educational objectives or occupational goals when mental conflicts, emotional maladjustments, and other types of personal problems are alleviated prior to or parallel with counseling and training. The plans and procedures of the Veterans Administration, therefore, are aimed at providing such thorough and complete counseling to each veteran claimant that he may be assisted in making the adjustments necessary for a useful life as a citizen. For this counseling service it is most important that competent, professionally trained, and otherwise well-qualified persons be employed.

For many veterans, guidance in the selection of an occupation or of an educational objective may be all that will be required. Other veterans may need assistance in handling

personal problems, in resolving mental conflicts, in attaining emotional stability, or in learning how to secure and hold employment. The Veterans Administration plans to furnish all types of counseling service which may be required in the case of an individual veteran. The veteran will also be guided in making intelligent use of other clinical and professional services, available to him through the Veterans Administration and other agencies, for the purpose of assisting him in making and maintaining the mental, emotional, and social adjustment essential to the attainment of his objectives. Each veteran is counseled in accordance with his needs as a person and educational and vocational guidance are not given without reference to the consideration of the other problems which affect the life of the individual.

The field organization of the Veterans Administration has been extended to carry out the general policies, plans, and procedures developed in the Central Office for counseling veterans. The program was started in 1943 by placing Vocational Advisers in the 53 regional offices in the different states. As the number of veteran claimants increase, Veterans Administration Guidance Centers are being established in a number of cooperating colleges, universities, and other educational institutions which have personnel qualified to render counseling services. This plan makes the counseling service accessible to veterans at points nearer their homes; it provides for the participation in this counseling service of a number of well-qualified people in colleges and universities who would not wish to sever their relationship with their educational institutions but who are able to contribute highly valuable service in the counseling of veterans; it affords veterans thorough and complete counseling in a suitable setting; and it enables the Veterans Administration to make arrangements, in advance, with educational institutions which, because they are going concerns, will be able to start counseling service to veterans on comparatively short notice when an increase in the number of claimants makes it desirable to have a larger number of Guidance Centers in any regional territory.

The fact that a veteran is requested to report for counseling

at a Veterans Administration Guidance Center located at any particular educational institution places no obligation upon him to take educational or training courses at that institution. Each Veterans Administration Guidance Center provides counseling service to any eligible veteran who may apply if he resides within the area to be served by the Guidance Center, and each veteran may take his training in any approved educational institution or training establishment where he can secure courses appropriate to the attainment of the educational or occupational objectives he has selected.

One of the most important advantages of the plan for establishing Guidance Centers in educational institutions is the provision it makes for a continuing supply of trained counselors, so that when the time comes to expand the service still further and to extend it from the nuclear points at the colleges and universities to the communities more remote from these it may be possible to make this expansion without reducing in any way the quality of the counseling service. Considering the probability that before long it may be necessary to provide counseling service at some 400 Guidance Centers at educational institutions and then to prepare to expand the service still more by supplying trained counseling personnel for additional offices of the Veterans Administration, it becomes apparent how important it is to make provision for a practical program of counselor training.

For this purpose the Guidance Center plan which locates the practical work of counseling in the educational institutions which provide counselor training is in many ways ideal. Such colleges and universities not only are able to train counselors through their usual classroom instruction, but they are also in a position to combine with this instruction observation of actual counseling procedure. As the new counselors develop they can be given closely supervised practice in counseling techniques. Thus the established Guidance Centers not only provide veterans with the services of well-qualified and experienced counselors, but they also serve to increase the number of trained counseling personnel.

At each Guidance Center the Veterans Administration will

have at least one Vocational Adviser and one Training Officer. The members of the faculty of the educational institution who are assigned to counsel veterans at Guidance Centers are called "Vocational Appraisers" in order to distinguish them, in the records, from the Veterans Administration Vocational Advisers.

The Veterans Administration Vocational Advisers are selected from lists of professional counseling personnel prepared by the U. S. Civil Service Commission, and the counselors who are employed by the educational institutions to render service to veterans in Veterans Administration Guidance Centers have similar professional qualifications. The success of the counseling program depends upon securing counselors with the requisite professional training, who are technically qualified in psychology and in tests and measurements, who are experienced in performing counseling functions and who can work effectively in an organization following prescribed procedures.

The counseling principles, methods, procedures, and techniques used in the Veterans Administration counseling program are contained in the "Manual of Advisement and Guidance" which is being printed by the Government Printing Office, and these will not be described here. Short, intensive training conferences covering the specific procedures and techniques which are described in the manual will be conducted at selected locations from time to time as part of the in-service training program for those entering the Veterans Administration counseling program in order to supplement the basic background of training and experience which the adviser of veterans should have acquired over a period of years.

The counseling program of the Veterans Administration both in the regional offices and in the Veterans Administration Guidance Centers adheres to the policy that counseling does not meet required standards if it merely informs veterans of occupational and training opportunities without making a careful survey and a thorough analysis of the individual's education, work history, abilities, aptitudes, interests, and personality traits. In such an individual survey and analysis, extensive use of objective tests is required in most cases. It has been the policy of the Veterans Administration to administer only

those tests which the counselor prescribes for each veteran, rather than to make use of a standard battery for all. Counselors in Regional Offices and in Veterans Administration Guidance Centers select from well-standardized tests those which they desire to use and their selections are submitted to Central Office for review and approval. This plan has enabled the Veterans Administration, while initiating the counseling program, to utilize the best skills of counselors whose experiences in the administration of tests have been quite different, without waiting to train them all in administering a specified list of tests. As the program goes forward and as experience in the testing of veterans in connection with counseling is studied and evaluated, it is probable that the tests which serve a specific purpose in a large number of cases will be more and more widely used and the selection will thus tend to become more standardized.

To take a long-time view, the Veterans Administration's policy of providing the services of professionally trained counselors for veterans should increase the number of persons who are trained for counseling service so that, after the present emergency is past, such counselors may be available to carry on more extensive student personnel programs in universities, colleges, and secondary schools. It should also enable community agencies to render more counseling services to adults. A greater understanding both of the values and of the limitations of tests and of other counseling techniques may be attained through experience in this program. Another result which may be expected will be an increased demand for counseling services and for professional training for counselors. This training will probably take a more practical form by including greater opportunity for clinical in-service training as well as for theoretical work in psychology and related fields.

# PERSONALITY TRAITS ASSOCIATED WITH ABILITIES. I. WITH INTELLIGENCE AND DRAWING ABILITY

RAYMOND B. CATTELL
University of Illinois

## I. *Conception of the Research Problem*

IT is customary to think of abilities as powers having an existence independent of other personality traits, i.e., of dynamic and temperament traits. In mathematical terms this is founded in the conception of unitary traits as factors (6). Since even different ability factors are mathematically independent of one another, it is not surprising that factors of different modality, e.g., temperament and ability traits, are still more confidently expected to be independent. Correspondingly, in clinical and general psychological terms the usual approach conceives of functionally independent traits or powers. Abilities, for example, are the tools of dynamic traits and may be used interchangeably by the same or different drives. Prediction of, say, the outcome of a son's antipathy to his father or a girl's overcompensatory concentration on school subjects rests first on an estimate of the strength of the drive but also on knowledge of the endowment in the various abilities which it may use.

The purpose of this paper is to show that the above analysis is only a first approximation to the truth. It proposes a more refined conceptualization and presents some new data, which, together with the data of an ensuing article (5), constitute a slight initial foundation for a factual edifice in this realm.

## II. *The General Nature of Ability–Personality Trait Connections*

Clinically the connection of abilities with dynamic traits is often quite striking. Inferiority overcompensations sometimes

produce astonishing performances—relative to the individual's I.Q.—either in all school subjects or in some obscure field in which the individual finds he can at once jump ahead of, and avoid outright competition with, his rivals (1). The writer has also known two or three "idiots savants" whose outstanding "trick" performance in computation or memorizing could be traced to some initial accidental display, of comparatively slight eminence, by which the individual discovered that he could step into the limelight from the drabness of institutional life.

Psychoanalysis provides abundant instances of special skills, perceptual and motor, developed, like symptoms, out of the unconscious drives, relentlessly seeking expression. The complementary phenomenon of *conscious* drives, as sentiments, shaping abilities is so commonly realized as scarcely to demand illustration. Many of the special abilities distinct from intelligence isolated by research, e.g., Thurstone's Primary Abilities, may prove to be environmentally, dynamically shaped patterns, from general ability being impressed by particular investments of time and energy in certain conventional patterns of skills. (This aspect has been discussed more fully elsewhere (4) in connection with *the theory of fluid and crystallized abilities.*) The influence of major, overt sentiments upon ability patterns probably reaches its widest expression in respect to the self-regarding sentiment and it may be, for example, that the lesser mechanical aptitude of girls, and even the failure to find a mechanical aptitude factor among measurements on girls, will in the end be traced, not to difference of natural capacities as such, but to differences of dynamic adaptation stereotypes of the self-regarding sentiment.

Naturally, causal connections can run in either direction, or in a "causal circle"; and from general psychological observation a clinician would confidently say that examples of these three theoretical possibilities exist with almost equal frequency. Interests produce discriminatory and motor abilities as discussed above. But the individual who finds himself endowed with certain good natural abilities is likely to enjoy exercising them, and, in a competitive world, to find the dynamic pattern of his self-regard increasingly shaped by these abilities. Consequently, the establishing of connections between abilities and

personality traits is only the first—though a highly necessary —stage in investigation, needing to be followed by exploration for causal connections.

### III. *Conditional and Wholistic Factors*

To dissect the present research problem into its ultimate roots one would need to ask whether the very distinction between ability, temperament, and dynamic traits may not be a chimera. Fortunately we have no need to enter upon so onerous an undertaking here, having carried it out elsewhere (7) and emerged with the conclusion that these common sense distinctions are real enough and capable of operational definition.

The psychologist familiar with correlation procedures is more likely to ask the very matter-of-fact question: "If abilities and non-cognitive personality traits sometimes correlate appreciably, how is it that the factors which have so far been discovered have been either pure ability factors, pure temperament factors, or pure dynamic factors?" The answer is, first, that there has been a tacit or unconscious conspiracy to maintain certain influences constant when giving tests, without mentioning—often without realizing—that such artificial conditions have been set up. We correlate ability tests given under conditions of quiet, of concentration, of common intention to do one's best. We correlate emotional responses in, say, nursery school children, observed under conditions in which cognitive abilities are not required in order to manifest emotion. That is to say, we always hold constant, in the measurement situation, the personality manifestations in which we are not interested.

Factors obtained under such test conditions we have named *conditional factors*. Since everyday life provides such constant conditions in a fair proportion of situations, conditional factors are by no means useless artificialities, when one comes to tasks of practical prediction.

Factors tend to be "pure" for a second reason, namely, that most investigators confine themselves to a deliberately narrow *sample* of the possible range of personality performances. They are looking for abilities, or even musical abilities only, or even ability to judge pitch alone. A very narrow range means less

communality, more unrelated specifics, fewer factors, and larger emphasis on the factor peculiar to that region.

The hypothesis can be put forward, therefore, that if one measured a collection of performances each under the naturally existing, varied conditions usually associated with the performances, and if one took an extremely wide sample of aspects of personality, many of the factors emerging therefrom would range at once over abilities, temperament traits, and dynamic traits. Factors obtained under such conditions we have called *wholistic factors*. Conditional factors will then appear as truncated forms of corresponding wholistic factors, cut down to their manifestations in a single modality.

General psychological considerations support such expectations. One can readily think of both *constitutional* and *environmental mold* source traits which should appear as wholistic factors. A single gene might be responsible for modifications in, say, certain parts of the midbrain, the cortex, the pituitary and the inner ear, causing simultaneous endowment in certain dynamic needs, some motor capacities, specific temperament traits, and particular auditory acuities. Similarly, particular school systems might frequently produce environment mold factors of covariation in certain acquired abilities, in forms of inhibition, and in specific interests and habits.

The exploration described below promised for the first time to reveal such wholistic factors if they exist; for it took a deliberately very wide array of personality aspects; in fact it sampled from what has been described in more detail elsewhere (2) as *the personality sphere*. Specifically it dealt with 35 clusters of traits, representative of all traits in the dictionary, rated for 208 adult males who had also been tested for intelligence, mechanical aptitude, drawing ability, verbal ability, mathematical ability, etc. The details have been described (3).

## IV. *Personality Associates of Intelligence*

The principal personality correlate of intelligence known to psychologists is moral character and it is with respect to this correlation that the vast majority of past data has been gathered. The correlation is established through studies of intelligence of delinquents, alcoholics, etc., contrasted with nor-

mal controls, and through correlations of intelligence with rated or measured moral character qualities in a "normal" population.

This whole field has been very systematically and competently surveyed and summarized by Chassell (11). The normal-delinquent and other dichotomous data on the one hand and the continuous data on the other, when expressed in corresponding correlation form, agree very well, as also do the data from ratings and those from such actual conduct studies as those of Hartshorne, May, and Maller (14). Chassell (11) summarizes the results from many researches and several thousands of subjects as follows: "Expressed in correlational terms, the obtained relation may therefore usually be expected to fall between .10 and .39, and the true relation to be under .50." Since most of these researches are on groups of restricted range the figure for the population as a whole would be higher. Chassell finally estimates that the correlation between intelligence and the moral character qualities of self-control, unselfishness, reliability, industry, loyalty, etc., when corrected for narrow sample and attenuation, is close to .60. There are clear indications that the correlation is somewhat higher for children than for adults and, curiously enough, higher for the correlation with intelligence than for correlation with school achievement. This last, together with the fact that measured intelligence apparently correlates as well as rated intelligence, reduces the weight of the possible criticism that in some rating (teacher-child) situations the "goodness of character" might be a spurious product of pleasing authority by doing good school work.

The findings of the present research may be viewed from two angles (1) as straight correlations viewed in terms of correlation clusters or surface traits and (2) in terms of factors or source traits.

1. Rated intelligence, in this group of adults, was found to take its place in two clusters, as follows:

| *Cluster Indexed as B3 in (2)* (Items in descending order of correlation with remaining items) | | | *Cluster Indexed as B4 in (2)* (Items in descending order of correlation with remaining items) | | |
|---|---|---|---|---|---|
| Original | v. | Banal | Intelligent | v. | |
| Constructive | v. | | Clear thinking | v. | Incoherent |
| Interests wide | v. | Interests narrow | Logical ability | v. | |
| | | | Given to reasoning | v. | |
| Independent | v. | Emotionally dependent | Clever | v. | |
| | | | Spatial-visual ability | v. | |
| Persevering | v. | Quitting | Mathematical ability | v. | |
| Intelligent | v. | Unintelligent | Analytical | v. | |

In these clusters every variable correlates with every other above .40 and the mean $r$ is about .75. Apart from various "cluster appendages" to these nuclei, provided by traits which correlate with some but not all of the items, these are the only two clusters into which intelligence enters. *B4* is essentially intelligence manifested as a cognitive ability surface trait. *B3*, which has also been found in the Sanford-Murray research, shows, on the other hand, intelligence amidst its closest personality associates.

2. In the factor analysis of the 35 clusters taken to represent all aspects of personality, one of the twelve factors—the second largest[1] in its contribution to the variance—appears to be a factor of general intelligence. Its loadings for all variables may be read in the published table (3). Here we shall discuss only traits with outstanding loadings. The six traits (each a cluster) with highest loadings are as follows:

*Loadings in Personality Factor B*[2]

| | | | | |
|---|---|---|---|---|
| (2) | .52 | (Intelligent) ..... | v. .... | (Stupid) |
| | | (Clear thinking) ..... | v. .... | (Incoherent, confused) |
| | | (Clever) ..... | v. .... | |
| (11) | .47 | (Persevering) ..... | v. .... | (Quitting) |
| | | (Painstaking) ..... | v. .... | (Slipshod) |
| | | (Conscientious) ..... | v. .... | (Conscienceless) |
| (4) | .43 | (Thoughtful) ..... | v. .... | (Unreflective) |
| | | (Deliberate) ..... | v. .... | (Impulsive) |
| | | (Austere) ..... | v. .... | (Profligate) |
| (28) | .42 | (Stable emotionally ..... | v. .... | (Changeable) |
| | | (Self-respecting) ..... | v. .... | |
| | | (Self-controlled) ..... | v. .... | (Unselfcontrolled) |
| (12) | .41 | (Intellectual) ..... | v. .... | |
| | | (Analytical) ..... | v. .... | (Unreflective) |
| | | (Wide interests) ..... | v. .... | (Narrow interests) |
| (3) | .41 | (Independent) ..... | v. .... | (Emotionally dependent) |
| | | (Reliable) ..... | v. .... | (Undependable) |
| | | (Mature) ..... | v. .... | (Emotionally immature, irresponsible) |

In the β rotation variable 28 falls slightly and 29 takes its place in the first six. 29 is

| | | |
|---|---|---|
| Alert ..... | v. ..... | Absent-minded |
| Energetic ..... | v. ..... | Languid |
| Quick ..... | v. ..... | Slow |

[1] It remains in this position, and substantially unchanged in character, in the alternative α and β factorizations which have been offered (10).

[2] This factor, which is well known as Spearman's "*g*" in ability studies, has been called "*B*" in the personality realm, to conform with the series of general personality factors and to express by alphabetical order its relative order of magnitude in that series.

The numbers at the left identify the cluster variables in the original list recorded in the factor analysis (3). When an actual intelligence test was correlated with the personality variables it confirmed the factor as being the ability factor of general mental capacity, outcropping in the personality realm, for it gave the following highest correlations, which agree substantially (5 of the 6 being the same) with those found in the factor analysis.

| | | $r$ |
|---|---|---|
| (2) | Intelligent, analytical, etc. | .40 |
| (12) | Intellectual interests, etc. | .31 |
| (11) | Strong-willed, conscientious | .30 |
| (29) | Psychophysically vigorous, alert | .29 |
| (3) | Wise, mature, polished | .26 |
| (35) | Smart, assertive | .24 |

One may next attempt to glean the evidence resident in past researches as to the correlation between intelligence and individual traits in the realm of character—as distinct from total character—to see if the above order is confirmed. Below we have taken the researches of Terman, Webb, and others reported by Chassell and averaged the $r$'s by a method roughly correcting the result for different samples and for attenuated coefficients. However, the method does not justify recording actual correlation magnitudes and we merely give the rank order of the traits, the $r$'s of which range from about .7 to .3.

| | | | |
|---|---|---|---|
| Cooperativeness (Highest) | (Average of | 7 | $r$'s) |
| Reliability, trustworthiness, responsibility | ( " | " 17 | " ) |
| Industriousness (mainly in school) | ( " | " 12 | " ) |
| Conscientiousness | ( " | " 10 | " ) |
| Sympathy | ( " | " 2 | " ) |
| Moral habits and ideals | ( " | " 6 | " ) |
| Unselfishness | ( " | " 2 | " ) |
| Sincerity (Lowest) | ( " | " 2 | " ) |

Until some remaining evidence is presented we shall defer discussion as to whether this order agrees with that of the loadings in our *B* factor. The remaining evidence of the nature of the *B* factor is to be obtained from looking into the traits which the factor *does not* load, i.e., those in the hyperplane of the vector. This is a necessary enquiry in the recognition of any factor. The *B* factor had presented some difficulties in rotating for simple structure because it produced only a relatively faint and uncertain hyperplane: there were relatively few variables

which it did not influence. (Only the *D* factor exceeded it in this width of influence.) The personality variables thus independent of the *B* factor (loaded 0.05 or less, in both α and β rotations) are:

| | | | |
|---|---|---|---|
| (5) | Hypochondriacal Self-deceiving | v. | Realistic |
| | Nervous and neurotic | | |
| (7) | Extra punitive | v. | |
| | Headstrong | v. | Gentle tempered |
| | Exhibitionist | v. | Self-effacing |
| (1) | Boastful | v. | Modest |
| | Assertive | v. | Submissive |
| | Conceited | v. | Self-critical |
| (9) | Thankless | v. | Grateful |
| | Hardhearted | v. | Softhearted |
| | Short-tempered | v. | Easygoing |
| (18) | High-strung | v. | Unexcitable |
| | Hurried | v. | Lethargic |
| | Vivacious | v. | " |
| (24) | Sadistic | v. | Not sadistic |
| | Suspicious | v. | Trustful |
| | Mulish | v. | Reasonable |
| (23) | Eloquent | v. | Inarticulate |
| | Flattering | v. | Natural, self-effacing |
| (32) | Optimistic | v. | Pessimistic |
| | Placid | v. | Worrying |

Inspection of the correlations of traits with the actual intelligence test confirms the independence of these traits with respect to general ability, and adds, within the 0.05 limit, the traits numbered 6, 14, 16, 19, 22, 26 and 33. Briefly these are: Softhearted, Antisocial schizoid, Active neurotic, Friendly frank, Aloof, Hypomanic emotional, and Introspective hurried, respectively.

If we study the factor loadings of the single variable "Intelligent" (No. 2), which has the highest loadings of any variable in the *factor* of general ability (*B*), we find that it has negligible loadings in any other factor. Its possibly significant loadings, all below .26, are in the factors *G, J,* and *K*—factors of character and education which, in the alternative β factorization, become absorbed in the *B* factor. In short, the direct personality associations of intelligence are wholly accounted for by the *B* factor loadings.

However, the *B* factor as a whole has some correlations with other factors, for we rotated for simple structure, even when attaining it meant leaving some factors somewhat out of orthogonal positions. Notably, this factor correlates positively

with *C* (.32) and, less certainly, with *D*, *G*, and *I*. When second-order factors (9, 17) are calculated from the correlations between the primary personality factors one, and possibly two, second-order factors are found, loading factors *B* and *C* positively and more highly than any others.

Adequate discussion of the meaning to be given to the novel concept of second-order factors is not possible here. (See 9 and 17.) The writer believes that the most likely explanation of the second-order factor loading *B* and *C* (Emotional Stability, inverse of General Emotionality) is that it is a social status factor, expressing the genetic adhesion (8) produced by assortative mating between intelligence, emotional stability, and other "success"-generating qualities that are intrinsically distinct.

So much for the sheer description of the connections between intelligence and personality structure. A fuller discussion of the possible interpretations and origins of these connections will be taken up towards the end of this article.

## V. *Personality Associates of Drawing Ability*

About the personality associates of creative artistic ability, notably about the alleged "artistic temperament," much has been written outside scientific psychology. One could wish that our data might throw light on this matter, but actually it is restricted to drawing ability and may or may not have reference to total artistic feeling and creativity. One hundred and twenty-eight subjects, selected from the 208 adult males in the total personality research for more uniformity of age and educational background, were asked to draw a man sitting in a chair reading a book. The drawings, all done within a 20-minute period, showed marked variations in resource, originality, and wit but were rated, by art teachers, for *artistic drawing ability*.[3] The reliability of the artistic ratings, as between the two judges, was .88.

Correlations with personality variables were carried out separately within each of the 8 groups of 16 within which the men were rated. The correlations agreed in sign in 6, 7, or all

[3] "Correctness of drawing, proportion, quality of line, expressiveness."

of the 8 groups for the following variables: 9 (–), 11 (+), 13 (+), 21 (+), 22 (+), 23 (+), 25 (+), 28 (–), 29 (+), 30 (+), 31 (+), 32 (+), 33 (+), 34 (+) (see variable list in 3). On a population of this size any correlation above about .22 is significant at the 1% level. The following mean correlations (for 8 groups of 16) are therefore worthy of especial note from among those listed above.

*Drawing Ability* with:

| | | | | |
|---|---|---|---|---|
| (34) | Labile | v. | Habit-bound | $r = .29$ |
| | Intuitive | v. | Logical | |
| | Careless of material things | v. | Thrifty | |
| (31) | Sociable | v. | Shy | $r = .29$ |
| | Responsive | v. | Aloof | |
| | Hearty | v. | Quiet | |
| (30) | Incontinent | v. | Inhibited | $r = .27$ |
| | Gluttonous | v. | Queasy | |
| | Curious | v. | Unenquiring | |
| (9) | Grateful | v. | Thankless | $r = .26$ |
| | Softhearted | v. | Hardhearted | |
| | Easygoing | v. | Short-tempered | |
| (33) | Tough | v. | Sensitive | $r = .25$ |
| | Lethargic | v. | Hurried | |
| | Talkative | v. | Introspective | |
| (29) | Alert | v. | Absent-minded | $r = .24$ |
| | Energetic-spirited | v. | Languid | |
| | Quick | v. | Slow | |
| (21) | Energetic-spirited | v. | Languid | $r = .24$ |
| | Self-confident | v. | Self-distrusting | |
| | Debonnaire | v. | ..... | |
| (22) | Responsive | v. | Aloof | $r = .22$ |
| | Genial | v. | Cold-hearted | |
| | Social interests | v. | Brooding | |

Because of the consistency of the correlation in all eight groups, despite its lowness, we suspect systematic connections also with (23) Exhibitionist, eloquent, flattering, (13) Easily jealous, self-pitying and (32) Optimistic.

The agreement of this syndrome with the popular stereotype of the earthy, gay, spirited, unstable, careless, Bohemian artist is very striking, yet it must be emphasized that at the time the ratings were made no one had any idea that the purpose was to correlate them with artistic ability. The correlations are low, but there is a reason additional to the simple statistical test of significance for believing that they are real. This is that the pattern of traits corresponds strikingly to a functional unity, or small group of functional unities, already found

in the factor analysis of personality (3). Seven of the above eleven variables occur among the small group of variables highly loaded with and used for defining the *F* factor of Surgency . . v. . . Melancholic Desurgency. Viewed more closely (see table of loadings, p. 88 of 3), this cluster of drawing ability personality associates is seen to be systematically produced by the *F* factor (Surgency) in collaboration with *H* (Charitable Rhathymia . . v. . . Obstructive Schizothymia) and to a lesser but perceptible extent with *E* (Dominance) and *J* (Vigorous, "Obsessional" Character). That is to say, if one picked out the variables high in these four factors, neglecting the influence of all other personality factors, and if one gave predominance to *F*, a slighter role to *H*, and a dash of *E* and *J*, the correlations between the members of this cluster would be fully accounted for. At this stage of research, however, no systematic, exact partialling out of the correlations has been attempted. The interpretation of this finding is discussed more speculatively below.

### VI. *Interpretation of Correlations with Intelligence*

The correlating of intelligence with a complete range of personality manifestations confirms the earlier impression of narrower experiments that it correlates almost exclusively with what may be called "character" qualities and that the correlations are of substantial magnitude. Moreover, among the character qualities themselves the traits persevering, conscientious, self-controlled, reliable, emotionally independent, industrious, etc., correlate somewhat more highly than unselfish, emotionally stable, sincere. It looks as if intelligence is *directly* more associated with character conceived in a narrow, self-conscious sense, and with respect to habits that are acquired later and through conscious ideals, rather than with basic emotional integration and goodness of character in the wider sense such as might result from the emotional adjustment derived from the upbringing of the first few years or from some relatively constitutional stability. The above character associates also demonstrably include breadth of interests, habits of reflective thought, and analytical habits of approach to problems.

The association with the above restricted realm of character qualities may be interpreted either as the functioning of a single factor *B* (β factorization), or due to three further factors, *G, J,* and *K,* which are themselves correlated with *B* as factors (α factorization). The pros and cons of the alternative rotations which give these two possible interpretations have been discussed elsewhere (10), and until further research is done a choice is impossible. However, no matter which of these analyses is accepted, the present writer considers that the general psychological knowledge in this field favors the hypothesis that the character correlations are due to better intelligence leading to better learning, which would be expected to show itself in conduct situations almost as much as in academic situations. That is, the character patterns are "environmental mold" traits, patterns of reward and punishment in the culture, which "take" better on a basis of good constitutional "*g*" endowment[4] than on poorer soil. At least during childhood it is an intelligent adaptation to develop good character. The converse hypothesis, that good character qualities lead to better performance in intelligence tests, can be given only a negligible role, in view, for example, of the negligible effect which conative variations of all kinds have been shown to have on intelligence test performance.

Although character qualities of the above restricted kind are *directly* to be connected with intelligence, the character qualities (for such they are commonly considered) of the *C* factor, which we may call deeper emotional integration or stability (see full description in 3), are connected only indirectly, by a second-order factor, saturating both *C* and *B,* and causing their appreciable correlation. We are inclined to think that most second-order factors will turn out to be best explained in causal terms as a common cause influencing both the correlated variables (i.e., as the third of the logically possible alternatives in explaining correlation).

[4] That these correlations run a little lower with adults may be due to the fact that years of trial and error experience compensate and reduce the learning gains due to intelligence alone. After all, much conduct learning is blind trial and error, even for the more intelligent. A radically different alternative would be that intelligent adults more quickly unlearn the moral habits taught them as children, but this would assume that intelligent people more frequently consider moral habits undesirable!

The *C* and *B* connection arises, we believe, from the effect of family and of social status (a) environmentally, in that more intelligent children, having more intelligent parents, will tend to get wiser handling of those emotional problems of the early years of childhood which determine the deeper emotional adjustment (c) and (b) genetically, by reason of what has been called the social status orientation of genetics patterns (9) showing itself in *socially conditioned genetic adhesions.* In this case, we presume, both high intelligence (positive *B* factor) and high emotional stability (positive *C* factor) are selected for social promotion and become genetically linked by assortative mating within social classes.

## VII. *Interpretation of Drawing Ability Correlations*

Previous studies of drawing ability[5] as an artistic performance, as distinct from Goodenough's ingenious use of it to measure intelligence (14), have on the whole failed to give us any clear picture of its ramifications as an ability or its connections with artistic abilities in general. Meier's survey (16) of the problem stresses that emotional and temperamental qualifications are as important as those connected with mere skills. Tiebout's (18) research, unfortunately on few cases, found no apparent superiority in various motor skills among those superior in drawing, but found superior power of observation and ability to retain visual impressions, for days or for months. Drep's (12) findings point the same way, showing better memory for visual form but not better motor ability, and indicating greater emotional sensitivity and more neurotic tendencies among abler artists than among others.

Artists familiar with the life of outstanding artists of the past and present were inclined to consider Van Gogh, Gauguin, Toulouse-Lautrec, or Whistler more typical than the staid Giotto and to impute higher general emotionality, impulsiveness, and ability to see everyday things freshly (through greater lability), and more imagination generally, to the true artist. One might add the consideration that psychotics have some-

[5] Many studies have unfortunately been on artistic ability as a whole, begging the question as to whether several distinct talents may not be involved.

times been known to paint for the first time after the onset of the psychosis, and that several instances are known, e.g., Churchill and Hitler, of vigorous minds turning to painting in times of political or professional failure. In short, the present writer would argue that any researcher in the future who makes statistical analysis of life *in situ,* i.e., of actual artists is likely to find the personality traits associated with artistic performance, as defined by these more intensive researches, amply illustrated.

Common observation also seems to indicate that there is a fairly powerful hereditary influence in drawing ability. On the bases of the factor associations found here we would argue contingently that it is largely a factor for the temperamental tendencies which is inherited and that the drawing ability *per se* is acquired on the basis of the interests which these generate.

The *F* factor of surgent temperament (associated in its extreme loadings with conversion hysteria), with its test manifestations of "fluency" and "imagination" which have long since been demonstrated to be part of it (1), seems to be the first condition for the development of drawing ability. The cyclothyme-like factor *H*—which, incidentally, is one of the two factors loading "aesthetic interests" in the original factor analysis—is the second factor having correlation with drawing ability and perhaps corresponds to the element of more passive sensitivity and appreciation in the interests which lead to drawing skill. The connection of cyclothyme temperament with appreciation of color and the visual arts was stressed by Kretschmer (15), who considered the high hereditary incidence of cyclothyme constitution in the Alpine-Mediterranean racial regions to account for their pre-eminence in production and appreciation of visual art. The indicated correlations of *E* (Dominance) and *J* (Vigorous, Obsessional Character) are too slight to justify discussion until further explored.

If clinical type observation, aided by the definiteness of factors engendered by factor analysis, may be admitted as a guide to further research, the present writer would suggest that the following associations are likely to become important when total, creative artistic ability, as distinct from drawing ability alone, is studied. First, observation suggests that the loadings

of the cyclothyme factor (*A* or *H*) will be found almost as important as those of Surgency (*F*). Secondly we can surely expect that factor *C*, in its negative loadings, will have some role. This factor connotes the instability, plasticity of thought and purpose, richness of emotionality and freshness of emotional approach, which the world evaluates one way and artists another. There is no hint of this factor in our results because we did not deal with creative artists but with soldiers selected for emotional stability and measured with respect to straight drawing ability. However, we suggest that it would be desirable to confirm the above personality factor associates of drawing ability before directing research to the more complex personality problem of artistic creativeness.

## VIII. *Summary*

*Intelligence.* Intelligence appears as a general factor (*B*) among personality traits, loading particularly *character* traits, and notably those good habits which may be consciously acquired. This factor correlates, however, to the extent of about .3, with a distinct factor (*C*) of emotional stability and integration, and together with other factors they yield a second-order factor, which may be the genetic adhesion of intelligence and temperamental (emotional) stability produced by social stratification. Past findings fit in well with these factorial interpretations.

*Drawing Ability* has significant correlations with eight out of the 35 surface traits used to represent the total personality sphere. These can be best explained as due to low positive correlations of drawing ability with Surgency (*F* factor) and Rhathymic Cyclothymia (*H* factor), and possibly slighter correlations with Dominance (*E* factor) and Vigorous Character (*J* factor). This personality pattern very distinctly resembles that observed in well-known artists, but it is suggested that total artistic ability, as distinct from artistic drawing ability alone, is also likely to involve General Emotionality (Negative *C* factor).

The writer wishes to express his thanks to Thelma Alper and Virginia Carvell for expert help in completing the drawing section of this research.

## REFERENCES


1. Cattell, R. B. *A Guide to Mental Testing.* London: University of London Press, 1936.
2. Cattell, R. B. "The Principal Trait Clusters for Describing Personality." *Psychological Bulletin,* XLII (1945), 129–161.
3. Cattell, R. B. "The Description of Personality. III. Principles and Findings in a Factor Analysis." *American Journal of Psychology,* LVIII (1945), 69–90.
4. Cattell, R. B. "The Measurement of Adult Intelligence." *Psychological Bulletin,* XL (1943), 153–193.
5. Cattell, R. B. "Personality Traits Associated with Abilities. II. With Verbal and Mathematical Ability." *Journal of Educational Psychology.* (In press.)
6. Cattell, R. B. "Personality Structure and Measurement. I. The Operational Determination of Trait Unities." *British Journal of Psychology.* (In press.)
7. Cattell, R. B. "Personality Structure and Measurement. II. The Determination and Utility of Trait Modality." *British Journal of Psychology.* (In press.)
8. Cattell, R. B. "The Cultural Functions of Social Stratification. I. Regarding the Genetic Bases of Society." *Journal of Social Psychology,* XXI (1945), 3–23.
9. Cattell, R. B. "Oblique, Second Order and Cooperative Factors in Personality Analysis." *British Journal of Educational Psychology.* (In press.)
10. Cattell, R. B. "Simple Structure in Relation to Alternative Factorizations of the Personality Sphere." *Journal of General Psychology.* (In press.)
11. Chassell, C. F. *The Relation between Morality and Intellect.* New York: Teachers College Bureau Publications, 1935.
12. Dreps, H. F. "The Psychological Capacities and Abilities of College Art Students of High and Low Standing." *Psychological Monographs,* XLV (1933), 134–146.
13. Goodenough, F. L. and Anderson, J. E. *Experimental Child Study.* New York: Appleton Century, 1931.
14. Hartshorne, H., May, M. A., and Maller, J. B. *Studies in Service and Self Control.* Vol. II of *Studies in the Nature of Character.* New York: Macmillan, 1930.
15. Kretschmer, E. *Physique and Character.* Berlin: Springer, 1929.
16. Meier, N. C. *Art in Human Affairs: an Introduction to the Psychology of Art.* New York: McGraw Hill, 1942.
17. Thurstone, L. L. "Second-Order Factors." *Psychometrika,* IX (1944), 71–100.
18. Tiebout, C. "The Psychological Functions Differentiating Artistically Superior from Artistically Inferior Children." *Psychological Monographs,* XLV (1933), 108–133.

# FACTOR ANALYSIS OF OCCUPATIONAL APTITUDE TESTS

Staff, Division of Occupational Analysis, War Manpower Commission

OCCUPATIONAL research projects have been conducted by the Division of Occupational Analysis since 1934, when it was known as the Occupational Research Program of the United States Employment Service. One of these projects has been the construction of aptitude tests. Batteries of aptitude tests have been validated for various occupations and are used in local offices of the Employment Service to aid interviewers in selecting the most satisfactory beginners for referral to jobs or training courses. In developing an aptitude-test battery for an occupation, a number of aptitude tests are "tried out" on employed workers (or trainees) for whom objective criterion data can be obtained, and the best combination of tests constitutes the battery.

The Division has recently been conducting a series of factor analysis studies to determine the important factors in its aptitude tests. Interest in the problem of the basic factors or fundamental aptitudes underlying the Division's occupational tests has been heightened by the recent emphasis on the counseling approach in vocational placement and guidance programs. From a practical point of view, the preparation of separate selection batteries for 20,000 different occupations is an unsurmountable task. However, if occupational proficiency can be expressed in terms of a few relatively independent aptitudes which, in various combinations, account for the differences in job aptitude requirements, then the use of tests for general counseling purposes becomes feasible. Several factor analysis studies have therefore been conducted in order to isolate the basic aptitudes measured by the Division's tests and to select a small number of measures of these factors for com-

bination into a counseling battery. It is planned to standardize this battery for a large number of occupations which will then be classified into groups on the basis of similarity of aptitude requirements.

### *The Experiment*

Several experimental batteries of tests were administered to a total of 2,156 persons for the factor analysis studies. The data are divided into nine experimental groups. The number of subjects, the number of tests, the geographical location, and the factors identified for each group are shown in Table 1. The smallest battery consisted of 15 tests; the largest, of 29. There was a great deal of overlapping of the tests among the several batteries; but, in all, some 59 tests were subjected to analysis. Some of the factors found are well known and are similar to those which have been discussed by Kelley, Thurstone, and other investigators. Others have not received attention heretofore.

In group 0, 19 tests were administered to 1,079 male applicants for defense training courses in Erie and Pittsburgh. The age of the subjects ranged from 17 to 39 years, with a mean of 23 years, and all had completed at least six years of education.

TABLE 1

*Description of the Experimental Groups*

| Group | Number of Subjects | Number of Tests* | Location | Factors |
|---|---|---|---|---|
| 0 | 1079 | 19 | Erie and Pittsburgh | O S P Q A T F M |
| 1 | 221 | 25 | Dallas and St. Louis | O V N S P Q A T F M |
| 2 | 99 | 29 | Sacramento | O N S P Q A T F M |
| 3 | 141 | 15 | West Virginia | O N S P Q T F M L |
| 4 | 138 | 25 | Philadelphia | O N S P Q T L |
| 5 | 275 | 27 | Cincinnati, Detroit, Cleveland, Toledo and Chicago | O V N S P Q A T F M |
| 6 | 98 | 28 | Chicago | O V N S P Q A T F M |
| 7 | 594 | 25 | Composite of Groups 1, 5 and 6 | O V N S P Q A T F M |
| 8 | 204 | 24 | Same as Group 2† | O N S P Q A T |

* Some of the correlational matrices were actually larger than here indicated, because of the inclusion of age and education as additional variables.

† This group includes all of Group 2 plus additional subjects from the same training course for whom data on five tests were not available.

In the remaining eight experimental groups, all of the subjects were males and most of them were trainees enrolled in Vocational Education National Defense Training courses. The age of the subjects ranged from 17 to 39 years, with a mean of 28 years. The mean number of years of education completed was 11, and ninety-nine per cent of the subjects had completed between 8 and 16 years. It is estimated that about five per cent of the sample were Negro, and the rest white.

### *General Description of the Tests*

A total of 59 different tests were employed in the several factor analysis groups. Fifty-four of these, 48 paper-and-pencil tests and six apparatus tests, were constructed by the Division. The other tests were the *O'Rourke Survey Test of Vocabulary* (*Form X4*), the *Revised Minnesota Paper Form Board* (Likert and Quasha), the *Minnesota Spatial Relations Test*, the *Minnesota Manual Dexterity Test—Placing*, and the *Minnesota Manual Dexterity Test—Turning*. In addition to the tests, age and education were included as variables.

In general, the tests constructed by the Division are speed tests, with time limits for the most part in the neighborhood of five minutes. The individual tests are homogeneous in content, since it was planned to combine them into batteries. The fact that they are intended for use in offices of the United States Employment Service, chiefly for industrial workers, is a central consideration. There was no need to construct many verbal or "intellectual" tests. Emphasis has been placed instead on development of tests of perceptual and spatial ability and of dexterity. It was intended to construct tests that appeared to have validity for occupations, but, on the other hand, were not so analogous to specific jobs as to impair the applicability of the tests for widespread use. All the tests are so constructed that they can be easily administered by personnel without extensive technical training.

### *The Factors*

Thurstone's methods (2, 3) were employed to extract the centroid factors from the correlational matrices and to rotate

them to a meaningful structure. For each group, a solution was first obtained which satisfied the criteria of simple structure. The effect of maximizing the number of zero loadings on as many factors as possible is equivalent to explaining test performance by a minimum number of factors. Simple structure is essentially the factor analysis analogue of the doctrine of parsimony.

It was discovered in each group that the first solutions had very nearly orthogonal structures. In one group, for instance, the largest correlation between factors was .099. The factors in an orthogonal structure are entirely independent and uncorrelated; when the factors are correlated among themselves the structure is said to be oblique. Since the structures were very nearly orthogonal, and inasmuch as the solutions were not so exact that different investigators would have obtained identical correlations between the factors, it was decided to impose an orthogonal structure on each group and the rotational process was continued until this was achieved. There is an important advantage to the final solutions so obtained: Comparisons of the results are rendered less ambiguous, in that reference can be made to factors which bear an identical relation to all other factors in each group.

The smallest number of common factors established in any group was seven, and the largest was ten. In all, eleven different common factors were found. Preliminary results from each group were applied to all the others, so that in a few groups more factors were determined than could be justified in any one of these if conducted independently. Consistent results were obtained from the several correlational matrices, in that the factors common to a related group of tests could always be demonstrated regardless of the composition of the remainder of the experimental battery. The loadings of a factor on a test for different groups varied to about the same extent as correlations for identical pairs of tests in the different groups.

Among the factors most readily established were the verbal ($V$), numerical ($N$), and spatial ($S$) factors. Although only three verbal tests appeared in the experimental batteries, the fact that their loadings ranged as high as .58 and did not fall

below .40 serves to identify this factor. One of these tests is the *O'Rourke Survey Test of Vocabulary*. Another is a proverbs test which measures the ability of the subject to understand proverbs. The third is a test of antonyms and synonyms.

The numerical factor (*N*) was found to be present in a wide variety of numerical tests: A test of arithmetic fundamentals (addition, subtraction, multiplication, and division); a specialized test of decimals and another of fractions; and reasoning problems, in which the subject solves verbal arithmetic problems. Two speed tests in one-digit arithmetic, one in addition, subtraction, and multiplication, and the other in addition only, are also classified as numerical tests.

The spatial factor (*S*) appears in about a dozen of the Division's paper-and-pencil tests, and was also found in the *Revised Minnesota Paper Form Board* and the *Minnesota Test of Spatial Relations*. The Division tests consist of two- and three-dimensional figures. Examples are surface development in which a pattern is matched with a three-dimensional object constructed from it, a picture test of the assembly of mechanical objects, another on fitting together abstract geometrical objects, and a test in selecting the mirror image of a line pattern. Thurstone remarks on the visual character of his spatial factor (4, 79–80) and the space tests in these studies may be similarly described. Kelley's description (1, 10), "manipulation of spatial relationships in so far as independent of differences in visual acuity," is also applicable. The existence of a perceptual factor tends further to limit the spatial factor. The large number of tests described above, all with substantial projections on this factor, make it easily identifiable as the same space factor found by other investigators.

One factor found in these studies presents difficulties in interpretation. This factor was found in each of the groups and is present in significant amount in about two dozen tests. The tests which have significant projections on this factor include all of the verbal tests, all of the numerical tests except the two speed tests of one-digit arithmetic, and almost all of the spatial tests. The factor was also present in a letter series test, a word memory test, and a perceptual relations test; this is in-

teresting because none of these tests have significant projections on either *V*, *N*, or *S*. It appears to have some of the properties of Spearman's *G*, but the two-factor theory has no place for group factors like *V*, *N*, or *S*. On the other hand, this factor has a wider significance and is more persistent than either Thurstone's *R* or *I* (4, 86–88; 5, 158–159). It appears to possess many of the properties that teachers, test examiners, and clinical psychologists would attribute to "intelligence." In Table 1 this factor has been designated, noncommittally, as Factor *O*.

A matter of interest, relating to current theories of general intelligence, is that this factor has been established in a sample of adults, ages 17 to 39. This tends to dispose of some theories that this factor could be established only among children, and that it amounts to a common maturational factor. The projection of age on this factor in this study is consistently negative, being about –.250. The loading of *V* on age, on the other hand, is consistently positive, about .320. From this one could possibly conclude that older individuals have greater facility in expressing themselves regarding familiar situations, and that younger individuals have greater facility in coping with new situations.

Among the remaining factors found are two perceptual factors, *P* and *Q*. *P* is present in a test on matching figures of various sizes and shapes, in a test on matching shaded figures, in a test on distinguishing figures which differ slightly in length, width, area, size of angle, or degree of curvature, and in many of the tests which also measure spatial ability. Altogether it appears to be present in over twenty tests. The factor *Q* is present in about nine tests, including a name comparison test and a number comparison test similar to those in the *Minnesota Test of Clerical Ability*, the arithmetical decimals test, a coding test, and a test on number copying. It is also found to some extent in several of the other tests which measure verbal or numerical ability. It may be further noted that the test of number comparison appears to measure *P* as well as *Q* and that the test of name comparison appears to measure *O* as well as *Q*. Another comparison test is one in which pairs of geo-

metrical figures are determined to be the same or different. Results are not conclusive but indicate that this test measures *P* and not *Q*.

The essential difference between the two perceptual factors is not entirely clear. For the most part the tests measuring *P* involve geometrical figures while the tests measuring *Q* involve words and numbers. Another way of interpreting the difference is to say that the tests which measure *Q* deal with materials emphasized in formal education, while the tests which measure *P* deal with materials to which little or no school time is devoted. The loading of *P* on age is about – .400; the loading of *Q* on age has not been positively established but is much nearer zero. Education has positive projections on both *P* and *Q*, but it is much higher on *Q*.

An aiming factor (*A*) was found in five paper-and-pencil adaptations of standard laboratory tests. In one test, the subject crosses the bars of a series of letter H's without touching the sides of the letter. In another test, the subject is required to make line strokes in squares. Apparently what is involved is accuracy or precision of movement. Whipple (6, 147–151) has applied the term "aiming" to various laboratory tests which measure accuracy of movement.

In the process of establishing orthogonal structures the speed factor (*T*) was found. It was discovered in one of the first groups that the factor *A* could not be established as a factor orthogonal to all the others. Leaving it in an oblique position would signify that the factor *A*, as then established, measured something also measured by the other factors. It was hypothesized that what was being measured in common was quickness or rate of movement, or speed. This was undoubtedly a general factor in a classical sense, since all the tests under study are speed tests. A general factor cannot be found directly by simple structure, and therefore the speed factor was established arbitrarily as a factor orthogonal to all the others. With the introduction of this factor, *A* was easily established in an orthogonal position without violating simple structure.

The factor *T* has less consistent loadings than any other factor, but approximately forty tests have significant projec-

tions on it in one or more groups. In general, it is easier to establish the existence of a phenomenon when it is sometimes present and sometimes absent than when it is always present. Related to this difficulty of establishing qualitatively the existence of this factor is undoubtedly the lack of consistent quantitative results. Among the tests with the highest loadings on *T*, however, are consistently those with moderate loadings on *A*. From this it follows that an increase in precision in the aiming tests results in a decrease in speed. A new speed test was constructed which demands practically no precision, and another is under construction demanding a large amount of precision. These tests will be included in future factor studies.

Two dexterity factors resulted from the apparatus tests. These have been designated *F* and *M* for finger dexterity and manual dexterity, respectively. An interesting finding was that the *Placing* and *Turning* parts of the *Minnesota Test of Manual Dexterity* had practically an identical factor composition, being almost pure tests of manual dexterity. Substantially the same finding obtains for Parts I and II of the *Peg Board Apparatus* developed by the Division. In Part I pegs are moved from one set of holes to another set of holes using both hands simultaneously, and in Part II the pegs are transferred from one hand to the other and replaced in the hole upside down. Another apparatus test involving fine assembly work measures *F*. However, disassembly of the same parts is a composite of *F* and *M*. *F* is significantly present in five tests and has moderate loadings on two others; and the existence of *M* is readily established from its presence in significant amount in five tests.

A factor *L* was tentatively established in two of the factorial studies, for four or five different tests. One of these is an analogies test consisting of line drawings and another is a verbal test in following directions. All of the tests with significant projections on this factor require the solution of problems by formal rational processes. This factor is a narrow reasoning factor and is being called the Logic factor. Age has a loading of about $-.350$ on *L*, and Education has a loading of $+.350$ on *L*.

It had been originally intended to include both a word

memory and a picture memory test in order to measure a possible memory factor. Unfortunately, only one memory test was administered, and no memory factor was found.

*Discussion*

Group factors have great significance for vocational counseling. This point may be developed briefly. If only a single general factor accounts for vocational success, all that one could do is establish a hierarchy of jobs, in order of general ability, as was done with the Army Alpha results after the last war. It would not be possible to distinguish, for instance, a potential lawyer from a potential engineer, assuming that engineers and lawyers possessed on the average an equal amount of general ability. If we assume, on the other hand, that engineering requires more ability than law, then it would follow that *every* individual would find law easier than engineering.

On the other hand, if one were to postulate that every task required its own separate specific ability, it would be impossible for the vocational counselor to assess all these abilities, and to predict all the vocational aptitudes the applicant might possess.

With group factors perhaps only a few hours of testing are required to sample the significant aspects of behavior. Since it is unlikely that each occupation requires a different set of aptitudes from that of every other occupation, those occupations with similar requirements could be grouped together into fields. Then, on the basis of a relatively small number of test scores, prediction could be made for a number of occupations.

## REFERENCES

1. Kelley, T. L. *Crossroads in the Mind of Man.* Stanford University: Stanford University Press, 1928.
2. Thurstone, L. L. *The Vectors of Mind.* Chicago: University of Chicago Press, 1935.
3. Thurstone, L. L. "A New Rotational Method in Factor Analysis." *Psychometrika,* III (1938), 199–218.
4. Thurstone, L. L. *Primary Mental Abilities.* Psychometric Monographs, 1938, No. 1.
5. Thurstone, L. L. "Experimental Study of Simple Structure." *Psychometrika,* V (1940), 153–168.
6. Whipple, G. M. *Manual of Mental and Physical Tests.* (2nd Ed.) Baltimore: Warwick and York, 1914. Part I.

# THE INTERESTS OF FOREST SERVICE MEN

EDWARD K. STRONG, JR.
*Stanford University*

THE top executive who has worked up from office boy has a profound belief that others could do likewise if they only would try hard enough. Is this true? Is willingness to work the major consideration or must abilities and interests also be taken into account? Suppose every nonsupervisory position was filled with a well-qualified man, would that insure a sufficient number for promotion to supervisory positions as vacancies occurred?

Many organizations, both public and private, maintain the policy of recruiting top management by promotions from below. Other organizations select men specifically for executive and administrative work and promote relatively few upwards from below. Personnel practices relative to selection and training reflect the policy of the organization involved. In far too many cases, however, no one in authority has ever definitely decided what the policy should be nor are the facts available upon which such a policy should be formulated.

Some data are given below which suggest that in the forest service the interests of district rangers and administrators differ so appreciably that it is questionable whether many of the former will or should be promoted to management. And it is furthermore questionable whether there are enough men in the lower bracket with the interests of administrators to supply the service with properly qualified men at the top.

In 1936, through the cooperation of the United States Forest Service, 410 members of that service filled out the *Vocational Interest Blank*. On the basis of these blanks an interest scale was developed to measure the degree to which a man has the interests of forest service men rather than the interests of men in general.

Such blanks may also be scored so as to reveal how similar

a man's interests are to the interests of men in 35 other occupations. Thus, it is possible to say that a particular ranger has more the interests of an engineer than of a forest service man and to say that a supervisor has not only the interests of a forest service man but also to a lesser degree the interests of a production manager, a personnel manager, a public administrator, and an office man. Such a supervisor might be expected to function somewhat differently from one who had an equally high forest service interest but whose secondary interests were those of an engineer, farmer, aviator, and policeman.

The data in this report are based upon the original group of 410 men plus about 50 additional cases of supervisors and regional and Washington administrators, obtained in 1941 in connection with a study of public administrators, financed by the Committee on Public Administration of the Social Science Research Council.

The data pertain to interests—what a man likes and also dislikes to do. The data do not include measures of general or specific abilities. Interests and abilities must both be considered before a final answer to the questions before us may be obtained. But what a man is interested in does play an important role, and it is this aspect of his behavior with which we are concerned here.

### *Scores of Forest Service Men on the Forest Service Interest Scale*

As stated above, the forest service interest scale is based upon the records of 410 men, whose average age in 1936 was 38.5 years. Approximately half were district rangers; all but three of the remainder were assistants to supervisors, assistant and associate supervisors, and supervisors. In terms of rank the criterion group averages just above that of district ranger. Forest service interests as measured by the scale represent the interests of men from district ranger to supervisor.

Table 1 gives the forest service interest scores of 430 men, distributed by age and rank in the service.[1]

[1] The classification of forest service men and the follow-up referred to below were made by Paul P. Pitchlynn, formerly assistant regional forester at San Francisco. The 410 blanks constituting the criterion group were also obtained largely through his efforts.

TABLE 1

*Number of Forest Service Men by Age and Rank; and Mean Forest Service Interest Score of Each Sub-Group*

| | | Age | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | |
| District Rangers | N | 37 | 38 | 40 | 30 | 24 | 13 | 7 | ... | 189 |
| | M | 53.2 | 52.5 | 51.5 | 47.5 | 46.0 | 47.9 | 38.2 | ... | 50.0 |
| Assistant to Supervisor | N | 11 | 3 | 3 | ... | ... | ... | 1 | ... | 18 |
| | M | 47.8 | 60.7 | 45.3 | ... | ... | ... | 39.0 | ... | 49.1 |
| Assistant Supervisor | N | 10 | 28 | 18 | 4 | 9 | 3 | 1 | ... | 73 |
| | M | 54.1 | 52.7 | 52.1 | 40.5 | 46.2 | 46.3 | 37.0 | ... | 50.8 |
| Associate Supervisor | N | 1 | 6 | 4 | 3 | ... | 1 | ... | ... | 15 |
| | M | 40.5 | 51.7 | 47.8 | 50.0 | ... | 54.0 | ... | ... | 49.7 |
| Supervisor | N | ... | 18 | 17 | 11 | 30 | 11 | 12 | ... | 99 |
| | M | ... | 47.5 | 47.5 | 51.1 | 48.3 | 48.9 | 41.3 | ... | 47.5 |
| P–6* | N | ... | ... | ... | 1 | 4 | 10 | 2 | 3 | 20 |
| | M | ... | ... | ... | 50.0 | 50.0 | 44.2 | 31.0 | 37.3 | 43.3 |
| P–7—P–8 | N | ... | ... | ... | 1 | 1 | 4 | 8 | 2 | 16 |
| | M | ... | ... | ... | 43.0 | 50.0 | 39.5 | 30.6 | 30.5 | 34.8 |
| Total | N | 59 | 93 | 82 | 50 | 68 | 42 | 31 | 5 | 430 |
| | M | 52.1 | 51.8 | 50.4 | 47.8 | 47.3 | 46.5 | 37.0 | 34.6 | 48.6† |

* P–6, P–7, and P–8 refer to grades of positions in the professional service of the Federal government, many of which entail a considerable amount of administrative responsibility, and with respective base salaries of $5600, $6500, and $8000 (as of the time of the administration of the test).

† Only 3 of the P–6 and none of the P–7 or P–8 personnel were included in the criterion group. If they were excluded and 13 cases, not so far classified, were included, the mean standard score of 410 criterion cases would be 50.

The data make clear that such scores decrease with age for all seven ranks in the service. The decrease in score amounts to only 4.8 from age 27.5 to 47.5 years, whereas it amounts to 15.7 from age 47.5 to 62.5 years.

Decrease with age in score on the forest service interest scale is a phenomenon strikingly peculiar to the profession. Data similar to those in Table 1 have been published for 29 occupations.[2] The average decrease in score for these 29 occupations, including forest service, amounts to 1.4 between the ages of 27 and 57 years, whereas the decrease is 15.7 for forest service men. There are appreciable decreases in score amounting to 10.5 for personnel managers, 7 for aviators, 5 for realtors and physicians and 4 for life insurance salesmen. The only occupation exhibiting increase of score on its own scale is that of minister, where

[2] E. K. Strong, Jr. *Vocational Interests of Men and Women.* Stanford University Press, 1943, Table 75.

the increase amounts to 6.9 score points. For most occupations age affects interest scores very little, but for some reason the reverse is the situation in the forest service group.

This phenomenon explains in part why a few successful forest service men score low in the forest service interest scale. They score low because they are older men and all older men average considerably below 50.[3]

The question naturally arises as to why forest interest scores decline with age among forest service men. Three possible explanations may be considered: first, such interests actually decline with age; second, there has been a change in the type of men entering the service and the younger men, constituting the larger proportion of the criterion group, have largely determined the norms for the group; and third, men with high scores leave the service in later life to a larger degree than men with lower scores.

The preponderance of data in our possession does not support the first explanation, that the kind of interests possessed by forest service men naturally decline with age. For example, when samples of 25-, 35-, and 45-year-old men drawn at random from the population are scored on the forest service interest scale the mean scores are, respectively, 28, 27, and 31, indicating a slight rise in score with age.[4]

How about the second explanation? In the early years of the forest service few men were college graduates, whereas in recent years many have been recruited from colleges, particularly from schools of forestry. Has this change in selection affected the interests possessed by forest service men?

The relationship of age to amount of education among 187 district rangers is as follows:

---

[3] The standard score of 50 is the average obtained by forest service men. For convenience scores are often expressed by the three letter ratings of A, B, and C. An A rating (scores of 45 to 75) means that the individual has the interests of men successfully engaged in the occupation; a C rating (scores below 30) means he does not have such interests; and a B rating (30 to 44) means he probably has those interests but we cannot be as sure of the fact as in the case of A ratings. Note that a C rating does not say a man is not interested in a particular occupation; it says he does not find interesting a whole range of activities which successful men in the occupation find interesting.

[4] *Ibid.*, p. 272.

| *Age* | *Number* | *Years of education* |
|---|---|---|
| 20–29 years ..... | 38 | 16.1 (college graduation) |
| 30–39 years ..... | 77 | 13.6 (nearly 2 years of college) |
| 40–49 years ..... | 54 | 10.2 (2 years of high school) |
| 50–59 years ..... | 18 | 9.0 (1 year of high school) |
| Total ..... | 187 | 12.7 ((2/3 year of college) |

On the basis of these data one might surmise that the decrease in forest service interest scores with age is attributable to lesser amounts of education. If, however, the age factor is partialled out we obtain the following interest score ratios for the six subgroups in terms of amount of education:

| | |
|---|---|
| 6 to 8 years, grammar school ..... | 83 |
| 1 to 2 years, high school ..... | 94 |
| 3 to 4 years, high school ..... | 107 |
| 1 to 2 years, college ..... | 95 |
| 3 to 4 years, college ..... | 101 |
| 1 to 2 years, graduate work ..... | 105 |

There is here no indication of increase in forest service interest score with increased amounts of education, except that those with only grammar school education do score lower than the remainder. But there are too few cases in this category to warrant emphasis upon the exception. Among supervisors there is no difference in forest service interest score between 24 college and 26 noncollege graduates. Apparently increase in amount of education has not affected forest service interest score. It is still possible that there have been other changes in selection besides education and that these undisclosed factors contribute to decrease in score with age but we have uncovered no proof of this.

The third explanation, that forest service scores decline with age because older men with high scores leave the service to a larger degree than those with low scores, can be established only by a long time follow-up of the men so far tested on the *Vocational Interest Blank*. Evidence presented below indicates that promotions, on the average, go to men with lower interest scores in mechanical pursuits and higher scores in general administrative interests. Such a hypothesis is tenable as far as the relationship of forest service and public administration interests is concerned, for they correlate only .21, which means that some men with high interest in one of these fields will have low interest in the other. It is therefore possible that some

men with high forest service interest scores have left the service when they failed to receive a promotion and that there have been enough such cases to explain the decrease in average scores with age.

*Occupational Interests of Forest Service Men*

So far we have considered only how far forest service men score on the forest service interest scale. Consider now how they score on other occupational interest scales.

Table 2 gives the percentage of rangers, supervisors, and administrators who rate A and B+ (scores of 40 to 75) in interest for 36 occupations. Such percentages indicate how many forest service men definitely have the interests of men in those occupations. For example, 88 per cent of rangers rate A and B+ on forest service interest in comparison with 82 per cent of supervisors, 55 per cent of P–6 officials, and 38 per cent of P–7 and P–8 administrators. As explained above, much of these differences is attributable to increasing age as we go from district ranger to P–7 and P–8 administrator, but we suspect not all.

Forest service men have in general the interests of skilled tradesmen (Group IV), particularly farmers; of production managers; of engineers; and of public administrators. Few of them have the interests of scientists (Groups I and II), office workers (Group VIII), salesmen (Group IX), lawyers-writers (Group X), and men engaged in social service (Group V).

Commenting upon Table 2, a former administrator of the service writes: "It explodes the old fiction that Forest Service men are scientists. True, some of us have had a little scientific training and some perhaps is a good thing, but we attempt to recruit scientists as rangers. Those we get go in one of three directions: some quit, some get transferred to research work, and some become frustrated and are no good to themselves or the Service."

There are some striking differences between the interests of district rangers near the bottom of the organization and the administrators at the top. Such differences are indicated fairly well in Table 2, especially in the case of those occupational

TABLE 2

*Percentage of Forest Service Men Rated A and B+ on Occupational Interests (i.e., Percentage Scoring 40 and Above)* *

| Group | Occupation | 49 District rangers | 44 Supervisors | 20 P–6 Washington & Regional | 16 P–7—P–8 Washington & Regional |
|---|---|---|---|---|---|
| III | Production Mgr. | 64 | 55 | 55 | 57 |
| IV | Mechanical Activities | | | | |
| | Forest Service | 88 | 82 | 55 | 38 |
| | Farmer | 86 | 59 | 50 | 6 |
| | Carpenter | 35 | 11 | 5 | 0 |
| | Aviator | 36 | 27 | 35 | 6 |
| | Printer | 28 | 13 | 15 | 0 |
| | Policeman | 34 | 23 | 35 | 6 |
| | Math. Sci. Teacher | 34 | 21 | 25 | 6 |
| II | Physical Sciences | | | | |
| | Engineer | 36 | 32 | 40 | 38 |
| | Chemist | 24 | 12 | 35 | 31 |
| | Mathematician | 0 | 0 | 5 | 6 |
| VIII | Office Activities | | | | |
| | Office Work | 26 | 16 | 30 | 13 |
| | Banker | 18 | 28 | 20 | 12 |
| | Accountant | 10 | 13 | 15 | 13 |
| | Purchasing Agent | 22 | 29 | 20 | 25 |
| IX | Sales Activities | | | | |
| | Realtor | 22 | 29 | 10 | 0 |
| | Sales Mgr. | 6 | 25 | 20 | 0 |
| | Life Insurance | 8 | 11 | 15 | 6 |
| XI | President | 16 | 18 | 20 | 32 |
| V | Social Service | | | | |
| | Public Administrator | 38 | 66 | 85 | 100 |
| | Personnel Mgr. | 14 | 23 | 50 | 38 |
| | Y. Physical Dir. | 14 | 4 | 15 | 6 |
| | Social Sci. Teacher | 8 | 11 | 30 | 19 |
| | City School Supt. | 2 | 11 | 25 | 19 |
| | Y.M.C.A. Secy. | 4 | 7 | 31 | 12 |
| | Minister | 0 | 5 | 5 | 6 |
| X | Linguistic Activities | | | | |
| | Author-Journalist | 8 | 14 | 10 | 31 |
| | Lawyer | 6 | 25 | 5 | 44 |
| | Advertising Man | 4 | 12 | 5 | 19 |
| VI | Musician | 6 | 0 | 0 | 0 |
| I | Biological Sciences | | | | |
| | Physician | 16 | 14 | 15 | 25 |
| | Architect | 8 | 2 | 10 | 25 |
| | Dentist | 10 | 9 | 5 | 0 |
| | Artist | 2 | 2 | 10 | 6 |
| | Psychologist | 0 | 0 | 10 | 12 |
| VII | Certified Public Accountant | 0 | 2 | 0 | 6 |

* The 49 rangers are a selection from 190 cases, so selected that the mean forest interest score and standard deviation of the 49 cases are practically the same as for the 190 cases. The 44 supervisors were similarly selected from 100 cases. The 20 P–6 and 16 P–7 and P–8 administrators are all the cases in our possession. The P–6 group contains 5 men stationed at Washington and 15 at regional offices. The P–7 and P–8 group contains 5 regional foresters and 11 men stationed at Washington.

interests in which a fair number of forest service men are interested. Such differences are, however, better shown by mean scores, for such take into account the men who score low as well as those who score high. Differences of four[5] or more in mean scores between district rangers and P–7—P–8 administrators are given in Tables 3 and 4, the former giving the occupational interests which decrease and the latter the occupational interests which increase, as one goes from district ranger to administrator. (In nearly every case the mean scores of supervisors and P–6 administrators lie very close to a plotted line connecting rangers and P–7—P–8 administrators. This fact adds support to the conclusions drawn from the data of district rangers and P–7—P–8 administrators alone.)

The interests which decrease (Table 3) are for the most part typical of mechanical activities, whereas the interests which increase (Table 4) are much more varied, being associated with administrative work, law-journalism, and social work.

The differences in interest scores in these two tables indicate that administrators differ in their interests from district rangers and, to a lesser degree, from supervisors. The differences imply that administrators are selected on a different basis from that used in the original selection of district rangers. This relationship will be found in most organizations, for administrators differ from the rank and file both in abilities and in interests.

One of the most notable differences in interests between district rangers and P–7—P–8 administrators is in the interests of public administrators. Only 38 per cent of district rangers rate A and B+ compared with 66 per cent of supervisors, 85 per cent of P–6 administrators and 100 per cent of P–7—P–8 administrators.

[5] Differences of 7 and 8 are statistically significant, i.e., have critical ratios of 3.0 and over, judging from a number of calculations; for example:

| | Forest Service | | Farmer | | Public Administrator | | Personnel Manager | |
|---|---|---|---|---|---|---|---|---|
| | Diff. | C.R. | Diff. | C.R. | Diff. | C.R. | Diff. | C.R. |
| Ranger vs. Super. | 5.0 | 2.3 | 5.0 | 2.8 | 4.5 | 2.4 | 5.5 | 2.4 |
| " P–6 | 8.0 | 2.9 | 7.0 | 3.0 | 11.5 | 4.8 | 12.5 | 4.3 |
| " P–7—P–8 | 16.0 | 6.1 | 14.5 | 7.6 | 12.0 | 5.9 | 11.0 | 3.7 |
| Super. vs. P–7—P–8 | 12.5 | 4.8 | 9.5 | 4.7 | 7.0 | 3.4 | 5.5 | 1.9 |

TABLE 3

*Occupational Interests in which Rangers Score at Least 4* More *than P–7—P–8 Administrators; also Differences in Scores of Younger and Older Rangers and Supervisors*

| | District rangers* | P–7—P–8 | Difference | Difference in scores of 30- and 50-year-old men | |
|---|---|---|---|---|---|
| | | | | Rangers | Supervisors |
| Forest Service | 52.5 | 36.5 | −16.0 | −10.3 | 2.0 |
| Farmer | 47.5 | 33.0 | −14.5 | 2.3 | 2.5 |
| Aviator | 38.0 | 25.0 | −13.0 | − 7.0 | −1.0 |
| Policeman | 37.5 | 23.5 | −14.0 | 7.5 | 4.0 |
| Carpenter | 36.0 | 14.5 | −21.5 | 7.7 | 9.5 |
| Printer | 35.5 | 24.0 | −11.5 | .7 | 5.0 |
| Math. Sci. Teacher | 35.0 | 29.0 | − 6.0 | − .5 | 3.5 |
| Purchasing Agent | 33.5 | 29.5 | − 4.0 | 2.5 | − .5 |
| Realtor | 33.5 | 29.5 | − 4.0 | 2.7 | −1.0 |
| Office Worker | 31.5 | 26.5 | − 5.0 | 3.7 | 3.5 |
| Dentist | 30.0 | 21.5 | − 8.5 | 1.1 | 2.0 |
| Y. Physical Director | 27.5 | 23.0 | − 4.5 | − 3.2 | − .5 |

* The rank order of occupations based on A & B+ ratings of rangers in Table 2 agrees very closely with rank order based on mean scores of rangers, some of which are given in this table and in Table 4. (The correlation between the two is .976.)

The public administrator scale is new. It is based on the interests of 518 men engaged in public administration. Included in the group are 46 supervisors and administrators of the forest service. The data in Table 2 are based on these 46

TABLE 4

*Occupational Interests in which Rangers Score at Least 4* Less *than P–7—P–8 Administrators; also Differences in Scores of Younger and Older Rangers and Supervisors*

| | District rangers | P–7—P–8 | Difference | Differences in scores of 30- and 50-year-old men | |
|---|---|---|---|---|---|
| | | | | Rangers | Supervisors |
| Public Administrator | 39.5 | 51.5 | 12.0 | −11.0 | −4.0 |
| President | 30.5 | 35.5 | 5.0 | .7 | −5.0 |
| Personnel Manager | 28.0 | 39.0 | 11.0 | −13.8 | −4.0 |
| Author-Journalist | 27.0 | 34.0 | 7.0 | − 2.8 | −3.0 |
| Lawyer | 25.5 | 37.5 | 12.0 | − 5.5 | −2.0 |
| Advertising Man | 25.0 | 33.0 | 8.0 | − 4.0 | −5.5 |
| City School Supt. | 21.5 | 34.0 | 12.5 | − 7.3 | − .5 |
| Y.M.C.A. Secy. | 21.5 | 25.5 | 4.0 | − 2.3 | 2.5 |
| Mathematician | 20.5 | 25.5 | 5.0 | − 3.3 | .5 |
| Minister | 16.0 | 22.5 | 6.5 | − 4.8 | 3.5 |
| C.P.A. | 18.0 | 30.0 | 12.0 | − 7.3 | −6.5 |
| Psychologist | 13.5 | 25.5 | 12.0 | − 9.8 | −3.0 |

men and in addition 34 supervisors-administrators and 49 rangers not included in the public administrator criterion group.

Scores on the public administrator scale correlate highest with scores on the scales of personnel manager (.75), city school administrator (.55), Y.M.C.A. secretary (.53), social science teacher (.53) and Y.M.C.A. physical director (.52). Since forest service administrators rate so very much higher than rangers on the public administrator scale it is to be expected that they will score higher on the social service occupations. The data in Table 2 do not show such increases except in the case of personnel manager. This is true because, with the exception of personnel manager, few forest service men score high on these occupational interest scales. Examination of mean scores shows, however, a steady rise in these interests from ranger to the P–7—P–8 level (see Table 4). But it must be emphasized that at the top level the mean score is 39 on the personnel manager scale, 34 on the city school administrator scale, 29 on the social science teacher scale, and 25 to 22 on the other three social service scales. Forest service men on the whole score low on the social service scales, scales which are related significantly to the interests of public administrators. This is true even of the forest service administrators.

### *Younger Men More Similar to P–7—P–8 Administrators than Older Men*

The fourth column in Tables 3 and 4 gives the difference in mean scores of 30(25–34)- and 50(45–54)-year-old rangers. Most of the differences are reversed from those between district rangers and P–7—P–8 administrators. That is, the older rangers differ from top administrators more than younger rangers. The same conclusion applies equally well to younger and older supervisors (last column in Tables 3 and 4).

The following correlations between interest profiles of groups of forest service men tell the same story.

| | |
|---|---|
| 30- vs. 50-year-old rangers | .861 |
| 30- vs. 50-year-old supervisors | .888 |
| 30-year-old rangers vs. P–7—P–8 | .402 |
| 50-year-old rangers vs. P–7—P–8 | .105 |
| 30-year-old supervisors vs. P–7—P–8 | .769 |
| 50-year-old supervisors vs. P–7—P–8 | .569 |
| P–6 vs. P–7—P–8 | .578 |

The interests of younger and older rangers are quite similar; the same is true of supervisors. Rangers' interests are not very similar to the interests of P–7—P–8 administrators, but the interests of the younger rangers are somewhat more similar than those of the older rangers. Supervisors' interests are much more akin to the interests of top administrators. Particularly is this true of the younger supervisors, whose interests are more similar to those of P–7—P–8 men than are the interests of P–6 administrators.[6]

The differences in interests of district rangers and administrators suggest that the man who is most typically a ranger is not likely to rise above the rank of supervisor and that promotions above the rank of supervisor are in terms of interests which are possessed by only a minority of district rangers. One former administrator comments, "Time after time I have seen our top rangers promoted only to lose interest and become mediocre, or at least no longer outstanding." A very real problem here as elsewhere is "how to determine in advance who will respond to promotion and who will not."

What we know about interests indicates that they are fairly permanent, especially among adults. There are cases where they have changed appreciably but such appear to be exceptions to the rule. If we employ the data on younger and older district rangers and supervisors as indicative of the changes attributable to increasing age, then such changes are in the wrong direction—older men are less like administrators than are younger men. If we assume that interests are fairly permanent and that the above changes are not attributable to increasing age, then it would appear, as suggested above, that some older men with distinctly ranger but not administrative interests are dropping out of the service when promotions are not forthcoming.[7]

[6] The difference of .297 between the two correlations of different age groups of district rangers has a critical ratio of 1.4, and the difference of .200 between the two groups of supervisors has a critical ratio of 2.3. Neither of these differences is statistically significant.

[7] The data suggest that district rangers are well selected for that work. But they do not possess the interests characteristic of administrators. Their pay is small considering their responsibilities. They should be rewarded by increased status and pay, not through promotion into a different type of work, but by keeping them on their present work for which they are suited and which they enjoy.

One cannot help wondering if the forest service is recruiting at the bottom enough men typical of top administrators to provide a good assortment from which in later years to select the leaders of the organization.

### *Recreational Interests*

In early days the forest service was concerned with the management of forest land involving lumbering, grazing, the construction of roads and buildings, and the fighting of fires. Then the public discovered the forests were a wonderful place for a vacation. This has focused greater emphasis upon the activity of handling people. Under the circumstances it is natural to ask, do the men selected for the original purposes of the service also possess the interests of men dealing with people? Is there any evidence that the younger men who have been selected in recent years have more social interests than the older men?

If we postulate that the occupations in Group V (see Table 2) typify the men who "handle others for their presumed good," then we can measure the extent to which forest service men possess social interests by noting their scores on the occupations in this group. Reference to Table 2 makes clear that few forest service men have such interests. The percentages are low for five of the six occupations and not at all high in the case of the sixth, i.e., personnel manager.

TABLE 5

*Interests of Recreational Administrators; also Differences in Scores of Recreation Men and Forest Service Men*

| Occupational interests | Mean score of recreation administrator | Differences in score between recreation men and | | | |
|---|---|---|---|---|---|
| | | District rangers | Superv. | P–6 | P–7—P–8 |
| Public Administrator .... | 48 | – 8 | – 4 | 3 | 4 |
| Personnel ..... | 44 | –16 | –10 | – 3 | – 5 |
| Social Science Teacher ... | 44 | –17 | –16 | –12 | –14 |
| City School Supt. ..... | 42 | –20 | –15 | –12 | – 8 |
| Y.M.C.A. Secy. ..... | 41 | –19 | –16 | –12 | –15 |
| Y. Physical Director .... | 40 | –12 | –12 | – 9 | –17 |
| Lawyer ..... | 36 | –10 | – 1 | – 5 | 2 |
| Math.-Science Teacher .. | 35 | 0 | – 4 | 0 | – 6 |
| Minister ..... | 35 | –19 | –17 | –12 | –12 |
| Average deviation ..... | .. | –13.4 | –10.5 | – 7.5 | – 9.2 |

If we postulate that the interests of public administrators engaged in recreational work typify the social interests which forest service men should now possess, we may then employ the interests of recreational men as a standard against which to check forest service interests. The nine occupational interests on which recreation administrators score 35 and higher are given in Table 5. It will be seen that six of the nine constitute the occupations in Group V referred to above. It will also be recalled that these interests correlate significantly with the interests of public administrators in general.

Forest service men score low on most of the interests listed in Table 5 (differences of 8 are statistically significant in almost every case). There is better agreement between the interests of recreation administrators and forest service men as we go from ranger to P–7—P–8 administrators, with the exception that P–6 administrators are slightly more similar to recreation administrators than are P–7—P–8 administrators.

When all 36 occupational interests are taken into account instead of only nine, as in Table 5, we have the following correlations between the interest profiles of recreation administrators and sub-groups of forest service men:

| | | |
|---|---|---|
| Recreation vs. | 30-year-old district rangers | −.01 |
| " " | 50-year-old district rangers | −.06 |
| " " | 30-year-old supervisors | .34 |
| " " | 50-year-old supervisors | .31 |
| " " | P–6 | .51 |
| " " | P–7—P–8 | .35 |

Evidently there is no relationship between the interests of recreation administrators and the interests of rangers, only a slight relationship in the case of supervisors and P–7—P–8 administrators, and some relationship in the case of P–6 administrators.

The above correlations between the interests of recreation administrators and forest service men may be compared with the following correlations between recreation administrators and eight other groups of public administrators:

| | | |
|---|---|---|
| Recreation vs. | Personnel men in public service | .80 |
| " " | Social insurance administrators | .75 |
| " " | Welfare administrators | .86 |
| " " | Publicity men | .29 |
| " " | Statistician | .16 |
| " " | Public health officials | −.12 |
| " " | Engineers | −.25 |
| " " | Chemists-Physicists | −.42 |

Seemingly, if the forest service is to handle the problem of recreation within the forests it must have men in the organization who understand such problems and genuinely enjoy dealing with them. Such interests are different from the interests of the typical forest man. Two ways of meeting the situation occur to us. First, men who possess both types of interest might be brought into the service. Second, men who are possessed of the recreational type of interest might be brought into the service to be specialists in this field. The former is a doubtful procedure because there are not many men who possess both sets of interests. (Forest service interests correlate with the interests of the occupations listed in Table 5 as follows: Public administrator, .21; Personnel manager, −.01; social science teacher, −.13; city school superintendent, −.25; Y.M.C.A. secretary, −.07; Y.M.C.A. physical director, .39; Lawyer, −.61; Math.-Science teacher .68; and Minister, .00.)

The second procedure would force a rearrangement by which recreational activities would at least be directed, if not carried on, by specialists. This would not be so convenient as the present procedure of having rangers carry on all types of activities. Whatever the organization, it certainly appears that there are not enough forest service men with the interests of recreation men to carry on such work enthusiastically.

Since the interests of administrators directing public recreational work correlate significantly with the interests of administrators in general, it is possible that adding such men to the forest service might result in increasing the number of younger men who would be selected later on for administrative work in the forest service.

### *Conclusion*

This report is restricted to the interests of forest service personnel. Abilities as well as interests must be considered before complete answers to the questions raised here can be obtained. The data suggest:

1. Forest service men have in general the interests of skilled tradesmen, particularly farmers; of production managers; of engineers; and of public administrators. Few of them have the

interests of scientists (Groups I and II), office workers (Group VIII), salesmen (Group IX), lawyers-writers (Group X), and men engaged in social service (Group V).

2. The older the man in the forest service, regardless of whether he is a district ranger or top administrator, the less he has the interests of district rangers-supervisors. Such decrease in interest score is apparently not attributable to increasing age, or shift in type of man entering the service, but to men with high scores leaving the service in greater numbers than men with low scores. Another bit of evidence in support of this explanation is that younger district rangers have interests somewhat more similar to those of top administrators than do older district rangers, the same thing being also true of supervisors.

3. District rangers differ in their interests from administrators in the forest service. The former have stronger mechanical interests and weaker interests associated with law-journalism, social work, and administrative work typified by city school superintendent, personnel managers, president of a business concern, and public administrator.

One cannot help but wonder if there are enough younger men in the service with interests of administrators so that the higher positions can be well filled twenty to thirty years hence.

4. Relatively few of the forest service personnel have the interests found among recreation administrators. But in recent years the forests have come to be used by millions of people for recreational purposes.

Should not some forest service personnel be specifically selected to handle the recreational facilities of the service instead of expecting this function to be taken care of by men already heavily loaded with other work and not possessing the interests of recreational enthusiasts?

5. It is likely that the rank and file of employees in most organizations differ in their interests from the executives.

Should men be selected to fill the lower positions with the hope that some will eventually fill the few higher positions as vacancies occur, or should some men be definitely selected for future placement in higher positions?

# A GROUP TESTING PROGRAM FOR THE MODERN SCHOOL

WARREN G. FINDLEY
New York State Education Department

IT is the fundamental thesis of this paper that the modern school has certain definite functions and characteristics which, in the light of the findings of educational psychology, call for a group testing program having equally definite characteristics.

## *Characteristics of the Modern School*

The modern school is characterized, first of all, by being charged with responsibility for educating all the children of all the people from school entrance until completion of grade 12 at age 18. At present some pupils do not complete grade 12 or remain in school until 18 years old, but increasing proportions do so, curriculum planning is based on this premise, and research is being pointed in the direction of finding why the schools fail to hold those that now leave without graduating. This trend toward the twelve-year common school has resulted in increasing individual differences in the pupil population in the upper grades with respect to interests and life goals, if not also with respect to intellectual ability.

A second important characteristic of the modern school, especially in the elementary grades, is the tendency to promote regularly practically all pupils. This is based on the psychological finding that the individual differences most significant in total development are social factors associated more closely with chronological age than with intellectual achievement. The practice of regular promotions finds further justification in the readily observed and experimentally tested deleterious effects of retardation on the morale of the retarded pupil. This trend in promotions policy has produced increased indi-

vidual differences in intellectual development in each school grade, so that teachers have been encouraged to "take each child where he is" and attempt to further his development as much as possible by adaptation of instruction to individual pupil needs.

A third characteristic of the modern school that bears on our problem has grown rapidly under the pressure of war findings that many high-school graduates have failed to maintain skill and ability at levels attained earlier. This condition has been found especially true of arithmetic. As a result the secondary school is now taking responsibility for maintaining and developing skills and abilities in basic areas like reading, arithmetic, and geography, previously accepted as mastered in the elementary school.

A fourth significant characteristic of the modern school is the emphasis on following and guiding total personality development of individual pupils cumulatively and constructively, and in line with this bringing to parents a clearer and less formal report of their children's progress and mastery of the work of the school. The old report card is giving way to interpretative communications, often including other matters besides scholastic achievement, calculated to provide the basis for better understanding and cooperation between the school and the home.

Other characteristics of the modern school might be noted—many of them significant—but the above will suffice for the problem under discussion.

### *Implications for the Testing Program*

The twelve-year common school—more than twelve years insofar as kindergarten and nursery school are added at the beginning—and the corollary acceptance by the secondary school of responsibility for maintaining and extending development of basic skills have made it possible to relax the attitude of frequent and grim testing in the early grades. There is no longer the necessity to view each year's work as possibly the last. With extended universal schooling it becomes possible to plan extended programs of instruction looking toward a

final pointing up at the end in more comprehensive examinations of the outcomes of the whole program. The best achievement at the end of the twelve years becomes the criterion and the justification for practices that lead to such success. In the pupil's whole school career the effort of the school properly goes to studying, guiding, and furthering his development toward ultimate achievement.

A program of testing progress along the main courses of the curriculum is called for if progress is to be guided. Such a program in New York State is called just that, a program of "progress testing" and "progress tests." *Reading Progress Tests* are designed to reflect progress in reading comprehension from grade 4 through grade 12. *Mathematics Progress Tests,* just issued, are similarly to reflect progress in mastering general mathematics from grade 4 to as high a grade as may involve general mathematical instruction. Similar tests are planned for the understandings, skills and appreciations in written English, science, health and safety, social studies, study skills, appreciation of literature, art appreciation, and music appreciation.

A further feature of this testing program derives from the tendency toward uniform promotions and the emphasis on interpreting individual achievement both within the school and to the parents. In such a program, testing must reflect progress along specific and directly improvable lines rather than in general terms of subjects. The *Reading Progress Tests* measure progress in three identifiable aspects of reading comprehension: ability to obtain detailed understanding, ability to discern central thoughts in passages, and ability to recognize meanings of words. The *Mathematics Progress Tests* measure progress in five aspects of problem solving: social information and quantitative concepts that are the basis for understanding the context of mathematical situations or problems, computational skill, ability to choose the operation or operations required to solve particular problems, ability to determine whether one has sufficient data to solve problems and to reject data irrelevant to the solutions, and finally ability to carry through the complete act of problem solving, using all the abilities separately measured in the other parts of the tests.

Many standardized tests of commercial publishers covering reading comprehension and English usage involve similar subparts devoted to particular skills in those areas. The *Iowa Every-Pupil Tests of Basic Skills,* especially *Test B: Work-Study Skills,* are so organized. *Test B* measures progress along five lines: reading maps, knowing sources of information, using the dictionary, using an index, and reading charts, graphs and tables.[1]

Tests thus subdivided provide the teacher with information about the relative strengths and weaknesses of the class as a whole and of individual pupils that permits planning of instruction to meet demonstrated needs. Summaries of achievement in this detail help administrators obtain more real understanding of the progress and mastery being accomplished under their general supervision. Profile charts showing the achievement of individuals and classes in the various skills bring out strengths and weaknesses especially clearly to both teacher and administrator. They also motivate pupil effort toward progress and provide a basis for intelligent understanding by parents of their children's status and progress. Under proper conditions, pupils may acquire satisfaction and hence motivation in plotting their own profiles from year to year in the areas tested.

A long step forward can be taken by establishing an annual fall testing program of progress tests. If promotions policy is to result in wide individual differences in every grade and promotion into a grade is no longer to be based merely on meeting minimum standards of achievement, fall testing in basic skills and areas of understanding will provide data essential to the teacher as she begins work with her new class. Testing at this

[1] The standardized testing approach described in preceding paragraphs applies well to testing progress in the mastery of skills. It has its limitations when applied to the content aspects of the social studies and science. Tests of content in these areas need to be adapted to varying local situations and programs and need to have a timeliness that standardized materials cannot maintain. It is true that much of the content of these areas that is important remains the same from year to year and from decade to decade. Standardized tests of these aspects, however, necessarily place a premium on permanent content at the expense of the timely and the local. Annually prepared school-wide, city-wide, or district-wide tests of the aspects neglected in standardized tests are essential if standardized tests of content are to be kept from exerting a restrictive influence on instruction through presenting a biased evaluation of content outcomes.

time gives up-to-the-minute evidence on what learning has survived the summer vacation to become the basis for further progress. For this very reason, such evidence is also more relevant to the administrator's interest in appraising the status and progress of his charges, than that which has commonly been obtained at the end of the school year.

Fall testing avoids the undesirable practice of cramming. Teachers will be less inclined to teach for specific tests. Testing at this time places a wholesome emphasis on collaboration between teacher and pupil to achieve goals of instruction in the year ahead.

If testing in the basic areas is done each year at the opening of school in the fall and the results are entered on cumulative records kept for each pupil, each teacher is provided not only with data concerning the current year's testing, but also with evidence of the previous progress of her new pupils. Achievement on this year's test takes on added significance when judged in the light of previous achievement and relevant notations about health, social, and emotional development, etc., that may be entered on such records. A pupil's seemingly mediocre achievement this year may represent significant improvement over his very poor achievement of earlier years, as a result of serious effort on his part and, perhaps, special assistance by his teacher. Such progress, which could not be inferred from the results of this year's testing alone, merits attention and encouragement.

What has been said above applies generally to the group testing of pupils from grade 4 through grade 12. In the primary grades, informal testing is generally to be preferred to standardized testing. In these grades children are still acquiring elementary skill in reading and the minimum vocabulary essential to dealing with standardized test materials. Printed standardized tests call for reading to such an extent as many times to make reading ability the main factor actually tested, regardless of the title and content of the particular tests, or they are devoted to testing only the mechanics of subjects, like computing in arithmetic or spelling in English. Until children have matured sufficiently to handle well-rounded tests, it

is better to omit formal group testing of the mechanical aspects of subjects. Until it is practicable in actual test situations to relate computation to problem-solving and spelling to effective writing, it is better not to risk drawing undue attention and effort to those aspects of the subjects that are easiest to test.

There are, however, some standardized tests that are useful in these grades. In the first grade or at the end of kindergarten, "readiness" tests may be helpful to teachers in judging pupils' readiness to undertake beginning study of the printed word and other work of the first grade. At the beginning of second and third grades carefully selected standardized reading tests will aid many teachers to appraise the general level of reading ability of each of their pupils, as in later grades the more detailed program of progress testing provides data on many counts.

A word regarding intelligence testing. Group intelligence testing—and group testing is most common—will add little to what is learned from a reading readiness test at primary levels. The unreliability of intelligence measures from group testing at these levels makes it wise to avoid such testing and the temptation the testing brings to enter a very doubtful I.Q. on the pupil's early record. In the intermediate grades the situation is quite different. Pupils are generally adaptable to group testing and results of intelligence testing provide a helpful clue in determining the approach to be made to a pupil having difficulty in his studies. One with a high I.Q. may be presumed to have prospects of considerable improvement if the specific source of difficulty in learning can be found; one with a low I.Q. may be expected to have considerable difficulty at least in the immediate future in most of his learning and should therefore be encouraged in his studies, but not exhorted to seek to attain tremendous progress. Annual administration of a group intelligence test at the beginning of grades 4, 5 and 6 should yield data immediately useful. The three separate testings will result in establishing a reliable measure for future as well as immediate reference. At the junior and senior high-school levels it may again be questioned whether anything useful in general school practice is accomplished by intelligence testing. What

has been learned of I.Q. in the intermediate grades and what may be ascertained by fall testing of skills in the upper grades provide all the necessary evidence of status and progress and a sufficient basis for prognosis of general or special success in advanced grades.

By way of summary, it may be said that the characteristics of the program of group testing recommended for the modern school are:

1. A twelve-grade testing program corresponding to the twelve-grade instructional program, with achievement by the end of grade 12 the ultimate measure of effectiveness of the total program.

2. Continuous testing of progress throughout the pupil's school career, recorded on cumulative records, as the basis for guiding his development toward ultimate achievement.

3. Use of tests yielding significant part scores related to specific, improvable skills, so that instruction may be related to demonstrated needs of individuals and whole classes.

4. Annual fall testing in each grade so that up-to-the-minute data may be available to all involved in guiding pupil development—teachers, administrators, parents, and the pupils themselves—at a time that is helpful and especially related to looking ahead.

5. Sparing use of formal tests below grade 4.

6. Use of group intelligence tests annually in the intermediate grades (grades 4 through 6).

Nothing has been said of diagnostic testing of a refined sort. Such testing is best conducted on an individualized basis, possibly following up leads suggested by the group testing program, but not as a part of the group testing program.

Group testing of interests and attitudes is not treated here. A group testing program in these areas is well justified, but falls outside the limits of this brief paper.

# REPLIES OF PSYCHOLOGISTS TO SEVERAL QUESTIONS ON THE PRACTICAL VALUE OF INTELLIGENCE TESTS

ARTHUR KORNHAUSER
Bureau of Applied Social Research, Columbia University

AN earlier report summarized the answers of a panel of mental test specialists to one part of a questionnaire on tests.[1] The present paper deals with the less technical questions which were asked at the same time.[2] Both reports are based on the replies of 79 psychologists. The questions were originally sent to 85 persons in this field who were chosen as representative of the most competent "experts" on mental measurement. The selection was based upon the pooled judgments of 6 advisory specialists who were consulted for the purpose. The names of the psychologists who participated in the poll are listed in connection with the report in the preceding number of this journal.[3]

*Question 1*

In your judgment, how well do intelligence tests meet the practical needs for classifying people as to general mental ability in the Army, in schools, and in industry?

| | *In the Army* | *In schools* | *In business and industry* |
|---|---|---|---|
| Extremely well; with a very small amount of error .... | 7% | 19% | 7% |
| Rather well, much better than is done without tests ..... | 81 | 78 | 60 |
| Not very well, but somewhat better than without tests .. | 12 | 3 | 33 |
| Not at all well; little or no better than without tests .. | 0 | 0 | 0 |
| | 100% | 100% | 100% |
| Number of cases ..... | 76 | 77 | 76 |
| No answer or not classifiable | 3 | 2 | 3 |

[1] This journal, Spring issue, 1945, pp. 3–15.

[2] The answers to these questions were obtained for the purpose of giving a popular summary of expert views on the matter to the public. This popularized report is published in *The American Magazine* for July 1945, as part of a new monthly "Poll of Experts" project.

[3] Loc. cit. Footnote p. 3.

An interesting result appears when the experts are divided according to whether or not they state that they have done test work in the armed forces, in schools, and in industry. The tests are rated a little higher in each field of application by those experts who have *not* worked in that particular field. While the relationships are slight and the numbers small, a consistent trend is present. The figures are as follows:

| | Has the expert done test work in the armed forces? | |
|---|---|---|
| | *Has* | *Has not* |
| The tests do "extremely well" in the Army | 2% | 13% |
| The tests do "not very well" in the Army | 14% | 10% |
| | (n = 43) | (n = 31) |

| | Has the expert done test work in schools? | |
|---|---|---|
| | *Has* | *Has not* |
| The tests do "extremely well" in schools | 19% | 25% |
| The tests do "not very well" in schools | 3% | 0% |
| | (n = 59) | (n = 16) |

| | Has the expert done test work in business? | |
|---|---|---|
| | *Has* | *Has not* |
| The tests do "extremely well" in business | 0% | 9% |
| The tests do "not very well" in business | 38% | 32% |
| | (n = 21) | (n = 53) |

Similar comparisons by age of the respondents show no tendency for the younger and older to differ in their ratings of test accomplishment. The same is true of clinical compared with non-clinical psychologists (classified according to their report of their own principal types of work).

When those who indicate "psychometrics" as a field of work are compared with others, a slight tendency is observed for the psychometricians to use the high and low rating categories a little more than do the others. Thus, the percentages of psychometricians saying "extremely well" with respect to the value of tests in the Army, in schools and in industry respectively are 9, 27, and 9; for non-psychometricians the corresponding figures are 3, 9, and 3. (The perfect 3 to 1 pattern is fortuitous!) For the low ratings of test accomplishment, the 2 sets of percentages are these: Psychometricians 14, 2, and 37; others 9, 3, and 27. This tendency may imply that the psychometricians had greater confidence in their own evaluation of test accomplishment than did the other psychologists.

*Illustrative Comments of the Respondents on Question 1*

Test administration in schools is usually slightly more careful than in the Army and in business and industry and consequently the results are slightly more reliable.

Better in Army and in business and industry because of greater heterogeneity of populations dealt with—less well in schools because of greater homogeneity. On the other hand, "general mental ability" is of more importance in schools and if adequate testing is done the job can be done extremely well in schools.

I would place the usefulness of a general classification test to schools and the Army above that to industry. This is because the problems in the last field are very frequently highly specific. Despite the greater complexity of Army problems I consider usefulness to the Army almost equal to that in the schools on the ground that tests classify the men much more accurately than other techniques under conditions where little time is available for processing.

The fields are still too general. For example in my present work in a large aircraft plant I find intelligence tests meet the practical needs for classifying accountants better than they do foremen and sub-foremen in the factory. Same would be true, I imagine, in the case of the Army.

Group tests below age 9 or grade 4, even when carefully administered, frequently yield unreliable determinations. At higher levels such tests do rather well in school.

The tests measure a type of general mental ability. The question is how important this rather abstract type of ability is in practical affairs. In school the importance is fairly clear, but in the Army or in business and industry the importance seems more questionable, and depends upon the specific job being studied.

The practical need is to classify people in terms of several specific *types* of mental ability; not in terms of *general* mental ability.

The concept of general mental ability has been largely discarded by mental test *experts* with experience in practical testing work outside of the school situation.

(Finally, there is one very long and critical reply from which two paragraphs are quoted, as follows:) Question I presupposes that there is some trait which one may properly call "general mental ability," and Question 2 likewise presupposes that there is a trait which one may properly call "mental ability," both traits being (a) unequivocally defined and (b) capable of being detected and measured *independently of the test that is employed to test them.* Are there such traits? If so, what operations determine them? If they exist, and if you define the appropriate operations, then both questions

have meaning. Only, the answer is to be found in an actuarial or contingency-table. It needs no census of experts' beliefs about what the proper answer would be. If there are no such traits as "general mental ability," or plain "mental ability," determinable independently of the tests which you have in mind, then every answer that you provide for will be factually false, since it presupposes a false assumption. . . .

If the tests are to be used, not to classify people as to "general mental ability" but rather to predict success in training, or the like, the answer is to be derived from the coefficient of validity, not by pooling personal impressions. Speaking by and large, these coefficients have run between (say) .3 and .5. For group prediction and rough screening they are valuable; for individual prediction almost worthless.[4]

*Question 2*

(a) How dependably do intelligence tests (the usual group tests for adults) measure the mental ability *of the individual person:* How safely can a person accept his test rating as a correct indication of where he will continue to stand in mental ability relative to others?

(b) What is your answer to the same question concerning the mental test rating *of the individual school child?*

| | *Ques.* (*a*) | *Ques.* (*b*) |
|---|---|---|
| Dependable measure of individual's ability; can be relied upon | 8% | 7% |
| Moderately dependable; seldom far wrong | 74 | 76 |
| Doubtfully dependable; often in error; not to be accepted without confirmation | 18 | 17 |
| Not at all dependable; likely to be misleading; should not be taken seriously | 0 | 0 |
| | 100% | 100% |
| Number of cases | 74 | 75 |
| No answer or not classifiable | 5 | 4 |

[4] A comment or two may be appropriate in defense of our questions about "intelligence tests" and measuring "general mental ability," in reply to these last 3 quotations (4 others expressed similar but less definite concern). The principal answer is that these concepts *are* widely used, both by psychologists and others, and that an immense amount of mental testing is aimed at measuring this general intellectual ability. Perhaps it would be better, as one respondent suggests, to employ the term "average mental ability" but prevalent usage seemed to warrant the assumption that test experts would recognize what is referred to in questions about "intelligence tests" and the practical classification of people on the basis of these measurements. There can be no doubt that practical efforts are frequently directed at predicting general alertness, adaptability, potentialities for learning, etc. The private convictions of particular psychologists regarding the futility of trying to measure these qualities is scarcely reason enough to justify discarding the concept and refusing to ask questions which pertain to it.

The other general criticism which is contained in the final quotation, to the effect that questions of the type we have asked can be answered simply by stating the size of correlation coefficients and standard errors, likewise seems untenable. There are many confusing statistical results on these matters which make it necessary for the "experts" to judge what are the typical or representative statistical findings. Moreover, even after particular figures are accepted, a problem remains as to the justifiable interpretation of the figures for practical purposes. It is precisely these final conclusions based upon the coefficients (plus the less quantitative evidence) which we were trying to ascertain.

Replies to these questions were compared for clinical and non-clinical psychologists and for psychometricians versus non-psychometricians. The clinical and the non-psychometric respondents are decidedly more harsh in their judgments regarding the dependability of the tests. On Question 2a the "doubtfully dependable" category (the lowest used by any rater) contained 31% of the clinical psychologists' answers and 39% of the non-psychometricians' as against 10% and 5% for the contrasting groups. Question 2b showed similar but less marked differences: 21% and 26% compared with 12% and 9%. On both questions the top rating category reveals slight differences of the same kind—that is, the psychometricians and non-clinicians tend to be a little more favorable in their estimates.

It seems probable, judging from some of the comments by the respondents, that the explanation for these differences may lie in the fact that a number of the measurement psychologists and non-clinicians answered in terms of simple short-run re-test reliability coefficients. The others tended to go beyond such figures and to consider also the varied and unequal conditions affecting individuals and their differing rates of growth over prolonged periods of time. The questions were intended to cover this larger problem rather than to refer to the narrower question of test reliability in a technical sense.

*Illustrative Comments of Respondents on Questions 2a and 2b*

Most group tests place the non-reader or slow reader at so great a disadvantage as to make results questionable until reading ability is determined.

But test results should be checked by reference to other sorts of evidence such as accomplishment in school and in occupation, and supplemented, when there are discrepancies or doubt, by tests administered individually.

Caution, of course, is necessary in the presence of unusually high or low ratings; verification with a more refined individual test is desirable in such instances.

Verbal tests are satisfactory for scholastic needs in school, but not for all other school needs. Likewise factors extraneous to test data may invalidate their (test data) classification value, e.g., low effort may negate high intelligence and vice versa. Wide diversity in values of different tests.

It would be better never to give anybody a single score called an "intelligence" score. Various aptitude scores should be used instead.

In my opinion incalculable damage has been done by the dogma of the "constancy of the I.Q." For example, in some school systems a child whose I.Q. at 9 years is 90 is not allowed to take algebra or a foreign language. Considering that the S.E. of prediction is about 9 points, one can see that some children are called dull at 9 who would be called normal or better at 11; while a very small proportion rated dull at one age would be rated near geniuses at another.

The use of group test results as predictors for individual cases is dangerous. High scores more meaningful than low; i.e., you deserve the score you get but maybe you should have more. Factors reducing scores (which are extraneous to the test itself) are far in excess of those raising scores.

In my estimation a good group intelligence test gives a moderately dependable estimate of the person's present functional level, but it does not predict his future standing as well as we like to think.

The above vote of confidence is predicated on the employment of the best tests; and one long enough to be reliable. Most short tests are not reliable.

My response here ("moderately dependable") refers to the "best" group tests. The widely used "self-administering" group tests are "doubtfully dependable."

The problem here is complicated by growth effects. A few children, destined to be normal, start growing late as if stunted. Repeated retesting, graphing the results, enables this to be detected.

I believe intelligence test ratings are a little more dependable for the older school child (high school) than they are for the younger ones.

Greater dependability with younger school children than with adults or older school children.

The younger the child, the greater the question.

Varies with the age of the child; very inaccurate in first years of life, and not dependable for adolescent years.

Whatever general mental ability exists is less subject to change in most cases after school-leaving than before.

I am more sure about this (dependability of person's test score) for school children than for adults because children are usually subjected to a similar school environment and abilities have a more equal chance to develop. With adults it is

possible to have differences occurring because of differences in environment—due to country of origin, work, etc.[5]

*Question 2c*

Are people with limited schooling rated correctly in comparison with persons who have more schooling, or do the former tend to be rated too low?

| | |
|---|---|
| Limited schooling not a source of error ..... | 0% |
| Source of slight error ..... | 38 |
| Source of considerable error ..... | 57 |
| Source of serious error ..... | 5 |
| | 100% |
| Number of cases ..... | 63 |
| No answer or not classifiable ..... | 16 |

Here, too, comparisons have been made in terms of whether the respondent checked "clinical work" or "psychometrics" as a principal part of his activities. Again, the psychometricians give slightly more favorable evaluations than the non-psychometricians. They give the rating of "slight error" in 39% of the answers as compared with 31% for the non-psychometricians; "serious error" is given by 3% as against 10% by the others. More of the psychometricians proportionately refuse to answer the question as asked, however (22% as compared with 12%). Clinical psychologists use both the upper and lower extreme ratings a little more than do non-clinicians ("slight error" 44% versus 30%, "serious error" 12% versus 3%) and are more reluctant to answer (24% versus 13% give no answers).

*Illustrative Comments of Respondents on Question 2c*

The comments under this question most frequently noted that the answer turns on the kind of test used or the nature of the test content (i.e., the problem is more serious with group tests and verbal tests than with individual and non-verbal ones). Twenty-four of the respondents made this point. The only other ideas expressed in many answers were that the reply depends upon how *extreme* the educational lack; whether due to absence of educational *opportunity* or not; and what the

[5] Interesting disagreements occur among these last six quotations. To some extent the apparent discrepancies doubtless stem from the ambiguity of the words "younger" and "older." Further inquiry would be required to find what other sources of disagreement are present.

nature of the cultural environment of the persons considered, apart from their schooling. Examples follow:

> The verbal group tests place a considerable premium upon formal education, particularly on the first few grades. Nonverbal tests tend to reduce this effect. Recency of education is also an important factor.
>
> Generally, the error is slight if exposure through at least the first six grades exists. (Another respondent says: Considerable error if less than 30 months of schooling.)
>
> Again depends on type of test employed and how cautiously interpreted. Also depends on cultural level of environment within which limited schooling obtains. Also depends on possibly related language handicap not due to lack of schooling in own language.
>
> This likelihood of error is particularly true for individuals who are dull. Unusually able individuals tend to rise above the limitations of schooling and to acquire such information by other means and from other sources.
>
> Generally speaking, mental ability determines amount of schooling. Bright individuals with limited schooling make high scores, whereas dull individuals who have been kept in school a long time make low scores.
>
> People with limited schooling rate lower than they will with more schooling. Whether this is an "error" depends on one's concept of intelligence.
>
> The larger the adequacy of the education, the more accurately does the test reflect the true capacity of the individual.
>
> In my opinion, the test tends to rate correctly those with limited schooling who have had ample opportunity for schooling; there is a source of considerable error where the group with limited schooling are those coming from communities where opportunities for schooling have been limited and where the limited schooling is the direct result of limited opportunities.
>
> Source of slight error for typical American communities; about 85 to 90% of children. No general statement is adequate; for perhaps 5% of our school children the error may be serious; and "considerable" for another 10% (especially in certain regions).

*Question 3*

(a) Do you find any serious misunderstandings or false expectations about intelligence tests (or "I.Q." tests) on the part of non-psychologists?

| | |
|---|---|
| Do | 92% |
| Do not | 4 |
| Doubtful or don't know | 4 |
| | 100% |
| Number of cases | 78 |
| No answer | 1 |

(b) If you do, what are the principal misunderstandings?

The types of misunderstandings mentioned by the respondents have been classified as shown in the following list. The items are arranged in descending order according to frequency of mention. The percentages are based on the number of respondents who named any classifiable misunderstanding, namely, 74 persons. Since the parts of a single answer were often classified under several headings, the percentages total more than 100.

*"Principal Misunderstandings"*

| | |
|---|---|
| Over-rating the test results; exaggerated belief in test validity, reliability, accuracy, constancy of I.Q., etc. | 58% |
| Belief that a test measures all aspects of ability; neglect of separate abilities; use of I.Q. for purposes for which not intended | 55% |
| Confusion in meanings of terms (I.Q., M.A., percentiles, intelligence, etc.); thinking of *any* test rating as an I.Q.; confusion of intelligence and information; wrong use of I.Q. applied to adults | 46% |
| Tendency to go to extremes in appraising tests; they are wonderful or they are worthless | 30% |
| Assumption that tests measure innate ability; that they are independent of environment | 22% |
| Other misinterpretations of what the tests measure or of their limitations | 19% |
| Failure to interpret scores in relation to norms or to think in comparative terms; misuse of norms | 16% |
| Under-rating the test results; exaggerated disbelief in test validity, reliability, etc. | 12% |
| Failure to recognize that some tests are better than others (group versus individual; limitations of verbal tests; etc.) | 14% |
| Too much credence given a single measurement, regardless of how and where test was administered | 10% |

# MEASUREMENT ABSTRACTS*

Altus, William D. "The Differential Validity and Difficulty of Subtests of the Wechsler Mental Ability Scale." *Psychological Bulletin,* XLII (1945), 238–249.

This is a report of a study made at an Army Special Training Center to determine the validity of the Army Wechsler subtests for use in counseling, guidance and prediction. The subjects were illiterate army trainees, and the criterion was "success-failure" in attaining a degree of literacy, comparable to the fourth-grade public school level. English-speaking subjects were given the verbal subtests of the Wechsler scale, while non-English-speaking subjects were given the performance subtests. Biserial correlations were computed and significant differences obtained. "Arithmetic Comprehension" and "Similarities" were highest on the verbal section, while "Digit Symbol" and "Series Completion" were high on the performance scale. *Francis Medland.*

---

Bailey, H. W. and Dallenbach, K. M. "A Study of Selective Procedures and Educational Achievement of ASTP Trainees Processed by the STAR Unit at the University of Illinois." *American Journal of Psychology,* LVIII (1945), 1–24.

The Specialized Training and Reassignment unit was organized to increase the efficiency of placement of candidates in Army Specialized Training Programs. Under its direction all candidates for ASTP took a basic battery of five tests: a) *Army General Classification* b) *Officer Candidate* c) *American Council on Education Psychological Examination* d) *Algebra* e) *Geometry.* These tests were measured against the "pass-fail" criterion for success in prediction. All showed "critical ratios" that were significant at the 5% level; the order of effectiveness of prediction from highest to lowest was: a) *Officer Candidate* b) *Algebra* c) *Geometry* d) *American Council on Education Psychological Examination* e) *Army General Classification. Francis Medland.*

---

Baldwin, E. F. and Smith, L. F. "The Performance of Adult Female Applicants for Factory Work in the Likert-Quasha Revision of the Minnesota Paper Form Board Test." *Journal of Applied Psychology,* XXVII (1944), 468–470.

The data for the table of norms presented in this paper were secured from the scores made by 975 women tested at the Hawk-Eye Works of the Eastman Kodak Co., Rochester, N. Y., ranging in education from seven years of schooling to graduation from college. Many nationalities and some Negroes were included. The testing administration and scoring were uniform and were done by the same person. The Hawk-Eye 16–25-year age group scored higher on all but one percentile level than the published norms. The Hawk-Eye 26–60 group was never higher than the 16–25, though on the published norms this condition existed at the 5th, 10th, and 15th percentiles. The authors believe their data may be more representative of adult female workers than the original norm group. *Elizabeth Bell.*

---

Bingham, W. E., Jr. "A Study of the Effect of the Presence of the Examiner upon Test Scores in Industrial Testing." *Journal of Applied Psychology,* XXVII (1944), 471–477.

Thirty-six men and 24 women averaging 21 years of age, all college students, were given equated forms of steadiness, typing, and addition tests in the examiner's presence and alone, the purpose being to check the validity of mechanical ability and personnel procedures where the job is a comparatively solitary one. Results

* Edited by Forrest A. Kingsbury.

indicated that the subjects were more efficient in each test when working alone, although in the addition test they completed more work with the examiner present. Introspection showed the examiner to be the stimulus factor for these effects. Further research is needed, according to the author. *Vernon S. Tracht.*

---

Blain, I. J. "The Rationale of Scientific Selection." *Occupational Psychology,* XIX (1945), 28–34.

This article supplements that of Cockett's, re-emphasizing some of its points and introducing others that had not been previously mentioned, all in line with the application of scientific techniques to the problems of personnel and job-placement in industry. Psychological tests are seen to give the employer a more objective, critical means of appraising a person's abilities than application forms, school reports, and interviews, although the latter are recognized as contributing information concerning personality, character, and temperament not as yet accurately measurable. *Vernon S. Tracht.*

---

Burton, Arthur and Joel, Walther. "Adult Norms for the Watson-Glaser Tests of Critical Thinking." *Journal of Psychology,* XIX (1945), 43–49.

The *Watson-Glaser Tests,* Battery I, were administered to 150 applicants for civil service positions. Ages ranged from 23 to 72 years, with the mean 39.1 and median 36.9. Subjects below the median age made significantly higher mean scores on all tests. The mean score of those with 2 or more college degrees was higher than those with 1 or no degree, but the difference was not statistically significant. Norms as a whole were higher than norms for college seniors. More extensive norms are needed, and a study of the validity of the tests in selection of professional and administrative personnel should be made. *Lorraine Bouthilet.*

---

Cockett, R. "The Rationale of Scientific Selection." *Occupational Psychology,* XIX (1945), 20–27.

Contrasting the autocratic with the democratic way by which a society may fully utilize the abilities of its individual members, the author describes the process of "scientific selection." This involves the organized method of discovering which persons possess the skills and capacities necessary for certain kinds of work, and of analyzing different jobs to determine what abilities are required for them. The experimental and statistical methods of securing valid and reliable measuring instruments for this purpose are briefly discussed and the personal or "human element" assessed. *Vernon S. Tracht.*

---

Goldstein, H. "A Malingering Key for Mental Tests." *Psychological Bulletin,* LXII, 104–118.

The malingering key is a scale composed of those items which proved most sensitive in differentiating between simulated malingering and genuine failure on the Army's *Visual Classification Test.* It is applied directly to the original test papers and yields a score based upon the number of discriminating easy items failed and the difficult items passed. The key is developed upon the hypothesis that morons and malingerers give test patterns differentiable because bona fide failures will fail the harder items, while the malingerers will tend to pass more of the hard items than the bona fide failures. In trials with thousands of cases, the key eliminated from seventy-five to ninety per cent of test failures as non-malingerers, thus enabling examiners to concentrate upon the remaining few and prove whether they were mentally adequate for service. *Elizabeth Bell.*

---

Jurgensen, C. E. "Report on the 'Classification Inventory,' a Personality Test for Industrial Use." *Journal of Applied Psychology,* XXVII (1944), 445–460.

Designed to avoid some of the faults of personality tests now used for predicting job success in industry, this inventory contains 245 items, comprising 45 groups of three items each and 55 paired comparison forms, and is intended to be scored and validated on jobs instead of personality traits. The author asserts that keys may be devised on the basis of traits for utilizing it in industrial, educational and clinical guidance work, but that occupational selection requires keys based on

specific jobs. Reliability and validity have been "satisfactory" in the two studies thus far completed, validation having been on groups comparable to those for whom the *Classification Inventory* is planned. *Vernon S. Tracht.*

---

Lowenfeld, V. "Tests for Visual and Haptical Aptitudes." *American Journal of Psychology,* LVIII (1945), 100–111.

Having determined that there are two groups of individuals with respect to their orientation: a) those who use their eyes as the main intermediaries for their sense impressions and b) those who, though with normal sight, depend upon touch and kinesthesis, the author has developed a battery of tests for use in job selection, where either the presence or absence of these abilities may be significant. On an experimental group of 1128 reactions, it was found that 47% were clearly visual, 23% were clearly haptical, and 30% were not distinguishable. Stress on this knowledge is important in job situations where the presence of one or the other aptitude would be a liability. *Francis Medland.*

---

McHugh, G. "Relationship between the Goodenough Drawing a Man Test and the 1937 Revision of the Stanford-Binet Test." *Journal of Educational Psychology,* XXXVI (1945), 119–124.

The author found an $r$ of .45 between M.A. scores and an $r$ of .41 between the I.Q. scores on the *Goodenough Drawing a Man* test and those on the *1937 Stanford Revision* (Forms L and M) of ninety kindergarten children in the public schools. These $r$'s, he believes, may be depressed because of the equal use of both forms of the *Binet.* A further study of the bi-serial $r$'s between individual scores of the two tests reveals that limiting the *Goodenough* to the nine items which have a biserial correlation coefficient of .30 or better with *Binet* I.Q. produces the best relationship between the two tests. Thus the number of Goodenough items used for kindergarten could be much reduced without losing reliability. *Elizabeth Bell.*

---

McNamara, W. J. and Weitzman, E. "The Effect of Choice Placement on the Difficulty of Multiple-Choice Questions." *Journal of Educational Psychology,* XXXVI (1945), 103–113.

It is generally believed that the "chance" element in "five-choice" and "four-choice" objective questions is "one-in-five" and "one-in-four," respectively. This study attempts to determine the possibility that placement of the correct choice in one of the four or five possible positions has a definite and measurable effect upon item difficulty. On a group of Naval Cadet subjects the next-to-the-last position, on both four- and five-choice items, had the greatest difficulty level. On four-choice items, the difficulty increased from the first through the third position. In five-choice items: the second and third positions were less difficult than the first, and the fifth position not significantly more difficult than the first. The data also indicate that the act of reading several incorrect choices has little or no effect upon the ability of the subject to select the correct choice in the series. *Francis Medland.*

---

Schmidt, H. O. and Billingslea, F. Y. "Test Profiles as a Diagnostic Aid: The Bernreuter Inventory." *Journal of Abnormal and Social Psychology,* XL (1945), 70–76.

These Army psychologists found that when subtests B1–N, B2–S and B4–D of the *Bernreuter Inventory* were considered in relationship to each other by statistical treatment, they were able to differentiate "standard normals" from "standard deviates" with an approximately 80% degree of certainty. Enlisted men of the Army Air Forces made up the 2 groups of subjects, 100 in the normal and 329 in the maladjusted, the latter having had psychiatric diagnosis by competent medical officers. While not indicating the direction of maladjustment, the inventory is believed by the authors to be a time-saving method of spotting deviate individuals. *Vernon S. Tracht.*

---

Slater, Patrick. "Scores of Different Types of Neurotics on Tests of Intelligence." *British Journal of Psychology,* XXXV (1945), 40–42.

This paper defends the "theory of overlapping group factors facilitating neurosis." Twenty-five men of each type of neurosis, obsessional, miscellaneous, anxiety

states and hysterias, were tested over a two-year period. The main tests were: *Progressive Matrices, Cattell IIa* and *IIB,* and the *Shipley Vocabulary Test.* Those with obsessional neuroses tended to be more intelligent than those with the other types. There was no proof that the others differed significantly; in other words, "neurotics are heterogeneous as regards intelligence." It is argued that if we find many common characteristics which differentiate neurotics from normal persons, we can expect also to find, when we isolate a particular characteristic, that the "neurotics are heterogeneous in its respect." *Elizabeth Bell.*

---

Staff, Psychological Section, Office of the Surgeon, Headquarters, AAF Training Command, Army Air Forces. "Psychological Activities in the Training Command of the Army Air Forces." *Psychological Bulletin,* XLII (1945), 37–54.

This is the seventh of a series of articles and is a report of the activities of the Section in the "application and correlation of the various tests used in the classification of aircrew members" and in the supervision and coordination of psychological research activities, including test development. Emphasis has been shifted from selection of aircrew trainees for success in training, which has been practically solved, to the selection of those who will make good combat officers. Data are being secured against which the tests may be validated. Efforts are being made to select men for the more specialized functions as members of lead crews, fighter pilots, and bomber pilots, and for the various types of gunnery training. There is also an extended search for proficiency criteria in training. *Elizabeth Bell.*

---

Wilson, G. M. and Burgess, Faye. "Construction Puzzle B as an Ability Test." *Journal of Educational Psychology,* XXXVI (1945), 53–60.

This is a discussion of the results of using *Construction Puzzle B,* a form-board test, as part of a battery of tests given in a war industry for classification purposes. It was observed that some subjects who were low on the form-board test were high on other tests. The need for special attention and interpretation, as well as the need for further study, in regard to this test is emphasized. *Lorraine Bouthilet.*